

Optimal Beamforming for Two-Way Multi-Antenna Relay Channel with Analogue Network Coding

Rui Zhang, *Member, IEEE*, Ying-Chang Liang, *Senior Member, IEEE*, Chin Choy Chai, *Member, IEEE*, and Shuguang Cui, *Member, IEEE*

Abstract—This paper studies the wireless *two-way relay channel* (TWRC), where two source nodes, S1 and S2, exchange information through an assisting relay node, R. It is assumed that R receives the sum signal from S1 and S2 in one time-slot, and then amplifies and forwards the received signal to both S1 and S2 in the next time-slot. By applying the principle of *analogue network coding* (ANC), each of S1 and S2 cancels the so-called “self-interference” in the received signal from R and then decodes the desired message. Assuming that S1 and S2 are each equipped with a single antenna and R with multi-antennas, this paper analyzes the *capacity region* of the ANC-based TWRC with linear processing (beamforming) at R. The capacity region contains all the achievable bidirectional rate-pairs of S1 and S2 under the given transmit power constraints at S1, S2, and R. We present the optimal relay beamforming structure as well as an efficient algorithm to compute the optimal beamforming matrix based on convex optimization techniques. Low-complexity suboptimal relay beamforming schemes are also presented, and their achievable rates are compared against the capacity with the optimal scheme.

Index Terms—Analogue network coding, beamforming, convex optimization, two-way relay channel.

I. INTRODUCTION

Network coding [1] is a new and promising design paradigm for modern communication networks: By allowing intermediate network nodes to mix the data or signals received from multiple links, as opposed to separating them by traditional approaches, network coding reduces the amount of transmissions in the network and thus improves the overall network throughput. Recently, there has been increasing attention from the research community to apply the principle of network coding in wireless communication networks. In fact, wireless network is the most natural setting to apply network coding due to the broadcast property of radio transmissions, i.e., a single transmission of one wireless terminal may successfully reach multiple neighboring terminals, without the need of dedicated links to these terminals as required in wireline networks. Furthermore, network coding can potentially be a very effective solution to the classical “interference problem” in wireless networks, since it transforms the traditional approach of avoiding or mitigating the interference among wireless terminals into a new methodology of *interference exploitation*.

Manuscript received August 1, 2008, revised February 16, 2009.

R. Zhang, Y.-C. Liang, and C. C. Chai are with the Institute for Infocomm Research, A*STAR, Singapore. (e-mails: {rzhang, ycliang, chaicc}@i2r.a-star.edu.sg)

S. Cui is with the Department of Electrical and Computer Engineering, Texas A&M University, Texas, USA. (e-mail: cui@ece.tamu.edu)

The *two-way relay channel* (TWRC) is one of the basic elements in decentralized/centralized wireless networks. The simplest TWRC consists of two source nodes, S1 and S2, which exchange information via a helping relay node, R. Traditionally, in order to avoid the interference at R, simultaneous transmission of S1 and S2 is unadvisable at the same frequency. Thus, in total four time-slots are usually required to accomplish one round of information exchange between S1 and S2 via R. However, by applying the idea of network coding, the authors in [2] proposed a method to reduce the number of required time-slots from four to three. In this method, S1 first sends to R during time-slot 1 the message s_1 consisting of bits $b_1(1), \dots, b_1(N)$ with N denoting the message length in bits, and then R decodes s_1 . During time-slot 2, S2 sends to R the message s_2 consisting of bits $b_2(1), \dots, b_2(N)$, and R decodes s_2 . In time-slot 3, R broadcasts to S1 and S2 a new message s_3 consisting of bits $b_3(1), \dots, b_3(N)$ obtained by bit-wise exclusive-or (XOR) operations over $b_1(n)$'s and $b_2(n)$'s, i.e., $b_3(n) = b_1(n) \oplus b_2(n), \forall n$. Since S1 knows $b_1(n)$'s, S1 can recover its desired message s_2 by first decoding s_3 and then obtaining $b_2(n)$'s as $b_1(n) \oplus b_3(n), \forall n$. Similarly, S2 can recover s_1 .

The principle of network coding has been further investigated for TWRC by exploiting various physical-layer relay operations [3], [4]. The scheme proposed in [3] is named as *analogue network coding* (ANC), while the one in [4] named as *physical-layer network coding* (PNC). For both ANC and PNC, the number of time-slots required for S1 and S2 to exchange one round of information is reduced from three [2] to two, by allowing S1 and S2 to transmit simultaneously to R during one time-slot and thereby combining the first two time-slots in [2] into one time-slot. ANC and PNC differ in their corresponding relay operations, which are amplify-and-forward (AF) and estimate-and-forward (EF), respectively. In ANC, R linearly amplifies the sum signal received from S1 and S2, and then broadcasts the resulting signal to S1 and S2. ANC is based upon an interesting observation that the signal collision at R during the first time-slot is in fact harmless, since such a collision can be resolved at S1 (S2) during the second time-slot by subtracting from its received signal the so-called *self-interference*, which is related to the previously transmitted message from S1 (S2) itself. In contrast to ANC, more sophisticated (nonlinear) operations than AF are required at R for PNC [4]-[7]. Instead of decoding messages s_1 from S1 and s_2 from S2 separately in two different time-slots like in [2], the EF method proposed in [4] estimates at R the bitwise XORs between $b_1(n)$'s and $b_2(n)$'s from

the mixed signal of S1 and S2, and re-encodes the decoded bits into a new broadcasting message s_3 ; each one of S1 and S2 then recovers the other's message by the same decoding method as that in [2]. Alternatively, it is possible to first deploy multiuser decoding at R to decode s_1 and s_2 separately, and then jointly encode s_1 and s_2 into a new broadcasting message s_3 ; given the side information on s_1 (s_2) at S1 (S2), S1 (S2) decodes s_2 (s_1). The above decode-and-forward (DF) relay operation for TWRC has been studied in [8], [9]. On the other hand, TWRC has also been studied in [10]-[13] from cooperative communication perspectives, with a major objective to compensate for the loss of spectral efficiency in the conventional *one-way relay channel* (OWRC) owing to the half-duplex constraint. Non-surprisingly, the solutions proposed therein are similar to those inspired by the principle of network coding.

Furthermore, TWRC has been studied jointly with other physical-layer transmission techniques based on, e.g., orthogonal-frequency-division-multiplexing (OFDM) [14], [15], and multiple transmit and/or multiple receive antennas [16]-[19], to further improve the bidirectional relay throughput. For the multi-antenna TWRC, the DF relay strategy was studied in [16], [17], the AF relay strategy or ANC was studied in [18], and the distributed space-time coding strategy for the relay was studied in [19]. In this paper, we focus on the AF-/ANC-based multi-antenna TWRC. Assuming that S1 and S2 each has a single antenna and R has M antennas, $M \geq 2$, we study the optimal design of linear processing (beamforming) at the relay to achieve the capacity region of AF-/ANC-based TWRC, which consists of all the achievable rate-pairs of S1 and S2 under the given transmit power constraints at S1, S2, and R. Our main goal is to provide insightful guidelines on the design of AF-based multi-antenna TWRC, which differs from the results for the conventional AF-based multi-antenna OWRC given in, e.g., [20]-[22]. The main results of this paper are summarized as follows:¹

- We derive the optimal beamforming structure at R, which achieves the capacity region of an ANC-based TWRC. The optimal structure reduces the number of complex-valued design variables in the relay beamforming matrix from M^2 to 4 when $M > 2$. Furthermore, by transforming the capacity region characterization problem into an equivalent relay power minimization problem under certain signal-to-noise-ratio (SNR) constraints at S1 and S2, we derive an efficient algorithm to compute the globally optimal beamforming matrix based on convex optimization techniques.
- Inspired by the optimal relay beamforming structure, we propose two low-complexity suboptimal beamforming schemes, based on the principle of “matched-filter (MF)” and “zero-forcing (ZF)”, respectively. We analyze their performances in terms of the achievable sum-rate in TWRC against the maximum sum-rate, or the sum-capacity, achieved by the optimal scheme. It is shown that the ZF-based relay beamforming with the objective

of suppressing the uplink (from S1 and S2 to R) and downlink (from R to S1 and S2) interferences at R may not be a good solution for the ANC-based TWRC, since these interferences are indeed self-interferences and thus can be later removed at S1 and S2. On the other hand, it is shown that the MF-based relay beamforming, which maximizes the signal power forwarded to S1 and S2, achieves the sum-rate close to the sum-capacity under various SNR and channel conditions.

The rest of this paper is organized as follows. Section II describes the TWRC model with ANC. Section III studies the capacity region of the ANC-based TWRC, derives the optimal structure for relay beamforming, and proposes an algorithm to compute the optimal beamforming matrix. Section IV presents the low-complexity suboptimal relay beamforming schemes. Section V analyzes the performances of both the optimal and suboptimal relay beamforming schemes in terms of the achievable sum-rate in TWRC. Section VI shows numerical results on the performances of the proposed schemes, in comparison with other existing schemes in the literature. Finally, Section VII concludes the paper.

Notation: Scalars are denoted by lower-case letters, e.g., x , and bold-face lower-case letters are used for vectors, e.g., \mathbf{x} , and bold-face upper-case letters for matrices, e.g., \mathbf{X} . In addition, $\text{tr}(\mathbf{S})$, $|\mathbf{S}|$, \mathbf{S}^{-1} , and $\mathbf{S}^{\frac{1}{2}}$ denote the trace, determinant, inverse, and square-root of a square matrix \mathbf{S} , respectively, and $\text{diag}(\mathbf{S}_1, \dots, \mathbf{S}_M)$ denotes a block-diagonal square matrix with $\mathbf{S}_1, \dots, \mathbf{S}_M$ as the diagonal square matrices. $\mathbf{S} \succeq 0$ means that \mathbf{S} is a positive semi-definite matrix [25]. For an arbitrary-size matrix \mathbf{M} , \mathbf{M}^T , \mathbf{M}^* , \mathbf{M}^H , and \mathbf{M}^\dagger denote the transpose, conjugate, conjugate transpose, and pseudo inverse of \mathbf{M} , respectively, $\mathbf{M}(i, j)$ denotes the (i, j) -th element of \mathbf{M} , and $\text{rank}(\mathbf{M})$ denotes the rank of \mathbf{M} . \mathbf{I} and $\mathbf{0}$ denote the identity matrix and the all-zero matrix, respectively. $\|\mathbf{x}\|$ denotes the Euclidean norm of a complex vector \mathbf{x} , while $|z|$ denotes the norm of a complex number z . $\mathbb{C}^{x \times y}$ denotes the space of $x \times y$ matrices with complex-valued elements. The distribution of a circular symmetric complex Gaussian (CSCG) random vector with mean \mathbf{x} and covariance matrix Σ is denoted by $\mathcal{CN}(\mathbf{x}, \Sigma)$, and \sim stands for “distributed as”.

II. SYSTEM MODEL

As shown in Fig. 1, we consider a TWRC consisting of two source nodes, S1 and S2, each with a single antenna and a relay node, R, equipped with M antennas, $M \geq 2$. All the channels involved are assumed to be flat-fading over a common narrow-band. It is assumed that the transmission protocol of TWRC uses two consecutive equal-duration time-slots for one round of information exchange between S1 and S2 via R. During the first time-slot, both S1 and S2 transmit concurrently to R, which linearly processes the received signal and then broadcasts the resulting signal to S1 and S2 during the second time-slot. It is also assumed that perfect synchronization has been established among S1, S2, and R prior to data transmission. The received baseband signal at R in the first time-slot is expressed as

$$\mathbf{y}_R(n) = \mathbf{h}_1\sqrt{p_1}s_1(n) + \mathbf{h}_2\sqrt{p_2}s_2(n) + \mathbf{z}_R(n) \quad (1)$$

¹Preliminary versions of this paper have been presented in [23], [24].

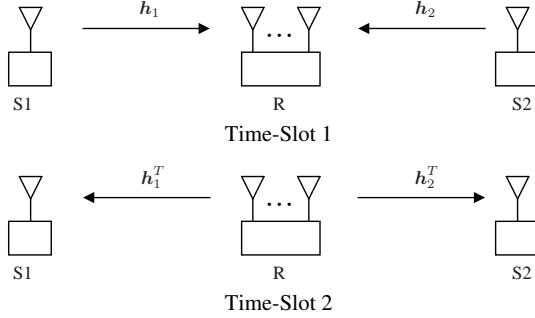


Fig. 1. The two-way multi-antenna relay channel.

where $\mathbf{y}_R(n) \in \mathbb{C}^{M \times 1}$ is the received signal vector at symbol index n , $n = 1, \dots, N$, with N denoting the total number of transmitted symbols during one time-slot; $\mathbf{h}_1 \in \mathbb{C}^{M \times 1}$ and $\mathbf{h}_2 \in \mathbb{C}^{M \times 1}$ represent the channel vectors from S1 to R and from S2 to R, respectively, which are assumed to be constant during the two time-slots; and $s_1(n)$ and $s_2(n)$ are the transmitted symbols from S1 and S2, respectively. Since in this paper we are interested in the information-theoretic limits of TWRC, it is assumed that the optimal Gaussian codebook is used at S1 and S2, and thus $s_1(n)$ and $s_2(n)$ are independent random variables both $\sim \mathcal{CN}(0, 1)$; p_1 and p_2 denote the transmit powers of S1 and S2, respectively; and $\mathbf{z}_R(n) \in \mathbb{C}^{M \times 1}$ is the receiver noise vector, independent over n , and without loss of generality (w.l.o.g.), it is assumed that $\mathbf{z}_R(n) \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$, $\forall n$. Upon receiving the mixed signal from S1 and S2, R processes it with AF relay operation, also known as *linear analogue relaying*, and then broadcasts the processed signal to S1 and S2 during the second time-slot. Mathematically, the linear processing (beamforming) operation at the relay can be concisely represented as

$$\mathbf{x}_R(n) = \mathbf{A}\mathbf{y}_R(n), \quad n = 1, \dots, N \quad (2)$$

where $\mathbf{x}_R(n) \in \mathbb{C}^{M \times 1}$ is the transmitted signal at R, and $\mathbf{A} \in \mathbb{C}^{M \times M}$ is the relay processing matrix.

Note that the transmit power of R can be shown equal to

$$\begin{aligned} p_R(\mathbf{A}) &= \text{tr}(\mathbf{x}_R(n)\mathbf{x}_R^H(n)) \\ &= \|\mathbf{A}\mathbf{h}_1\|^2 p_1 + \|\mathbf{A}\mathbf{h}_2\|^2 p_2 + \text{tr}(\mathbf{A}\mathbf{A}^H). \end{aligned} \quad (3)$$

We can assume w.l.o.g. that channel reciprocity holds for TWRC during uplink and downlink transmissions, i.e., the channels from R to S1 and S2 during the second time-slot are given as \mathbf{h}_1^T and \mathbf{h}_2^T , respectively.² Thus, the received signals at S1 can be written as

$$\begin{aligned} y_1(n) &= \mathbf{h}_1^T \mathbf{x}_R(n) + z_1(n) \\ &= \mathbf{h}_1^T \mathbf{A} \mathbf{h}_1 \sqrt{p_1} s_1(n) + \mathbf{h}_1^T \mathbf{A} \mathbf{h}_2 \sqrt{p_2} s_2(n) \\ &\quad + \mathbf{h}_1^T \mathbf{A} \mathbf{z}_R(n) + z_1(n) \end{aligned} \quad (4)$$

for $n = 1, \dots, N$, where $z_1(n)$'s are the independent receiver noise samples at S1, and it is assumed that $z_1(n) \sim \mathcal{CN}(0, 1)$, $\forall n$. Note that on the right-hand side (RHS) of (4),

²This assumption is made merely for the purpose of exposition, and the results developed in this paper hold similarly for the more general case with independent uplink and downlink channels.

the first term is the self-interference of S1, while the second term contains the desired message from S2. Assuming that both $\mathbf{h}_1^T \mathbf{A} \mathbf{h}_1$ and $\mathbf{h}_1^T \mathbf{A} \mathbf{h}_2$ are perfectly known at S1 via training-based channel estimation [26] prior to data transmission, S1 can first subtract its self-interference from $y_1(n)$ and then coherently demodulate $s_2(n)$. The above practice is known as *analogue network coding* (ANC) [3]. From (4), subtracting the self-interference from $y_1(n)$ yields

$$\tilde{y}_1(n) = \tilde{h}_{21} \sqrt{p_2} s_2(n) + \tilde{z}_1(n), \quad n = 1, \dots, N \quad (5)$$

where $\tilde{h}_{21} = \mathbf{h}_1^T \mathbf{A} \mathbf{h}_2$, and $\tilde{z}_1(n) \sim \mathcal{CN}(0, \|\mathbf{A}^H \mathbf{h}_1^*\|^2 + 1)$. From (5), for a given \mathbf{A} , the maximum achievable rate (in bits/complex dimension) for the end-to-end link from S2 to S1 via R, denoted by r_{21} , satisfies

$$r_{21} \leq \frac{1}{2} \log_2 \left(1 + \frac{|\mathbf{h}_1^T \mathbf{A} \mathbf{h}_2|^2 p_2}{\|\mathbf{A}^H \mathbf{h}_1^*\|^2 + 1} \right) \quad (6)$$

where the factor $\frac{1}{2}$ is due to the use of two orthogonal time-slots for relaying. Similarly, it can be shown that the maximum achievable rate r_{12} for the link from S1 to S2 via R satisfies

$$r_{12} \leq \frac{1}{2} \log_2 \left(1 + \frac{|\mathbf{h}_2^T \mathbf{A} \mathbf{h}_1|^2 p_1}{\|\mathbf{A}^H \mathbf{h}_2^*\|^2 + 1} \right). \quad (7)$$

Next, we define the capacity region of ANC-based TWRC, $\mathcal{C}(P_1, P_2, P_R)$, subject to transmit power constraints at S1, S2, and R, denoted by P_1 , P_2 , and P_R , respectively. First, for a fixed pair of p_1 and p_2 , $p_1 \leq P_1$ and $p_2 \leq P_2$, we define the achievable rate region for S1 and S2 as

$$\mathcal{R}(p_1, p_2, P_R) \triangleq \bigcup_{\mathbf{A}: p_R(\mathbf{A}) \leq P_R} \{(r_{21}, r_{12}) : (6), (7)\}. \quad (8)$$

Then, $\mathcal{C}(P_1, P_2, P_R)$ is defined as

$$\mathcal{C}(P_1, P_2, P_R) \triangleq \bigcup_{(p_1, p_2): p_1 \leq P_1, p_2 \leq P_2} \mathcal{R}(p_1, p_2, P_R). \quad (9)$$

Note that in (9), $\mathcal{C}(P_1, P_2, P_R)$ can be obtained by taking the union over all the achievable rate regions, $\mathcal{R}(p_1, p_2, P_R)$'s, corresponding to different feasible pairs of p_1 and p_2 . Thus, for the rest of this paper, we focus our study on characterization of $\mathcal{R}(p_1, p_2, P_R)$ for some fixed p_1 and p_2 . Also note from (8) that the relay beamforming matrix \mathbf{A} plays the role of realizing different rate tradeoffs between r_{21} and r_{12} on the boundary of $\mathcal{R}(p_1, p_2, P_R)$.

For the convenience of later analysis in this paper, we express (4) into an equivalent matrix form as follows, combined with $y_2(n)$'s and $z_2(n)$'s defined for S2 similarly as for S1.

$$\begin{bmatrix} y_2(n) \\ y_1(n) \end{bmatrix} = \mathbf{H}_{\text{UL}} \mathbf{A} \mathbf{H}_{\text{UL}}^T \begin{bmatrix} \sqrt{p_1} s_1(n) \\ \sqrt{p_2} s_2(n) \end{bmatrix} + \mathbf{H}_{\text{DL}} \mathbf{A} \mathbf{z}_R(n) + \begin{bmatrix} z_2(n) \\ z_1(n) \end{bmatrix} \quad (10)$$

where $\mathbf{H}_{\text{UL}} = [\mathbf{h}_1, \mathbf{h}_2] \in \mathbb{C}^{M \times 2}$ and $\mathbf{H}_{\text{DL}} = [\mathbf{h}_2, \mathbf{h}_1]^T \in \mathbb{C}^{2 \times M}$ denote the uplink (UL) and downlink (DL) channel matrices, respectively. Note that $\mathbf{H}_{\text{DL}} = \mathbf{F} \mathbf{H}_{\text{UL}}^T$, where $\mathbf{F} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$.

III. CAPACITY REGION CHARACTERIZATION

In this section, we study the capacity region of ANC-based TWRC by characterizing $\mathcal{R}(p_1, p_2, P_R)$ defined in (8) for a given set of p_1 , p_2 , and P_R . First, we derive the optimal relay beamforming structure for \mathbf{A} that attains the boundary rate-pairs of $\mathcal{R}(p_1, p_2, P_R)$. It is shown that with the optimal beamforming structure, the number of unknown complex-valued variables to be sought in \mathbf{A} is reduced from M^2 to 4 when $M > 2$. Then, we formulate the optimization problem and present an efficient algorithm to compute the optimal \mathbf{A} 's to achieve different boundary rate-pairs of $\mathcal{R}(p_1, p_2, P_R)$.

A. Optimal Relay Beamforming Structure

Let the singular-value-decomposition (SVD) of \mathbf{H}_{UL} be expressed as

$$\mathbf{H}_{\text{UL}} = \mathbf{U}\Sigma\mathbf{V}^H \quad (11)$$

where $\mathbf{U} \in \mathbb{C}^{M \times 2}$, $\Sigma = \text{diag}(\sigma_1, \sigma_2)$ with $\sigma_1 \geq \sigma_2 \geq 0$, and $\mathbf{V} \in \mathbb{C}^{2 \times 2}$. It thus follows that $\mathbf{H}_{\text{DL}} = \mathbf{F}\mathbf{V}^*\Sigma\mathbf{U}^T$. We then have the following theorem.

Theorem 3.1: The optimal relay beamforming matrix, \mathbf{A} , that attains a boundary rate-pair of $\mathcal{R}(p_1, p_2, P_R)$ defined in (8) has the following structure:

$$\mathbf{A} = \mathbf{U}^* \mathbf{B} \mathbf{U}^H \quad (12)$$

where $\mathbf{B} \in \mathbb{C}^{2 \times 2}$ is an unknown matrix.

Proof: Please refer to Appendix I. \blacksquare

Remark 3.1: In the conventional AF-based multi-antenna OWRC, the optimal beamforming structure at relay to maximize the end-to-end channel capacity has been studied in, e.g., [20], [21]. Applying the results therein to the OWRC with S2 transmitting to S1 via R yields the optimal \mathbf{A} to maximize r_{21} in (6) as $\mathbf{A}_{21} = c_{21}\mathbf{h}_1^*\mathbf{h}_2^H$, where c_{21} is a constant related to P_R . Similarly, the optimal \mathbf{A} to maximize r_{12} in (7) for the OWRC from S1 to S2 via R is in the form of $\mathbf{A}_{12} = c_{12}\mathbf{h}_2^*\mathbf{h}_1^H$. It then follows that \mathbf{A}_{21} differs from \mathbf{A}_{12} unless $\mathbf{h}_1 = v\mathbf{h}_2$ for some constant v , i.e., \mathbf{h}_1 and \mathbf{h}_2 are parallel. Therefore, relay beamforming designs for the OWRC with *separate unidirectional* transmissions in general can not be applied to the TWRC with *simultaneous bidirectional* transmissions. As observed from Theorem 3.1, the optimal relay beamforming matrix for TWRC lies in the space spanned by both \mathbf{h}_1 and \mathbf{h}_2 .

Let $\mathbf{g}_1 = \mathbf{U}^H\mathbf{h}_1 \in \mathbb{C}^{2 \times 1}$ and $\mathbf{g}_2 = \mathbf{U}^H\mathbf{h}_2 \in \mathbb{C}^{2 \times 1}$ be the “effective” channels from S1 to R and from S2 to R, respectively, by applying the optimal structure of \mathbf{A} given in (12). Similarly, \mathbf{g}_1^T and \mathbf{g}_2^T become the effective channels from R to S1 and S2, respectively. $\mathcal{R}(p_1, p_2, P_R)$ in (8) can then be equivalently re-expressed as

$$\begin{aligned} \bigcup_{\mathbf{B}: p_R(\mathbf{B}) \leq P_R} \left\{ (r_{21}, r_{12}) : r_{21} \leq \frac{1}{2} \log_2 \left(1 + \frac{|\mathbf{g}_1^T \mathbf{B} \mathbf{g}_2|^2 p_2}{\|\mathbf{B}^H \mathbf{g}_1^*\|^2 + 1} \right), \right. \\ \left. r_{12} \leq \frac{1}{2} \log_2 \left(1 + \frac{|\mathbf{g}_2^T \mathbf{B} \mathbf{g}_1|^2 p_1}{\|\mathbf{B}^H \mathbf{g}_2^*\|^2 + 1} \right) \right\} \end{aligned} \quad (13)$$

where $p_R(\mathbf{B}) = \|\mathbf{B}\mathbf{g}_1\|^2 p_1 + \|\mathbf{B}\mathbf{g}_2\|^2 p_2 + \text{tr}(\mathbf{B}\mathbf{B}^H)$. Note that the not-yet-determined parameter in (13) is \mathbf{B} . Since \mathbf{B} has 4 complex-valued variables as compared to M^2 in \mathbf{A} ,

the complexity for searching the optimal \mathbf{B} corresponding to a particular boundary rate-pair of $\mathcal{R}(p_1, p_2, P_R)$ is reduced when $M > 2$. Using Theorem 3.1 and (13), optimal structures of \mathbf{A} can be further simplified in the following two special cases, which are *Case I*: $\mathbf{h}_1 \perp \mathbf{h}_2$, i.e., $\mathbf{h}_1^H \mathbf{h}_2 = 0$; and *Case II*: $\mathbf{h}_1 \parallel \mathbf{h}_2$, i.e., $\mathbf{h}_1 = v\mathbf{h}_2$ with v being a constant.

Lemma 3.1: In the case of $\mathbf{h}_1 \perp \mathbf{h}_2$, the optimal structure of \mathbf{A} is in the form of $\mathbf{A} = \mathbf{U}^* \begin{bmatrix} 0 & c \\ d & 0 \end{bmatrix} \mathbf{U}^H$, with $c \geq 0$ and $d \geq 0$.

Proof: Please refer to Appendix II. \blacksquare

Lemma 3.2: In the case of $\mathbf{h}_1 \parallel \mathbf{h}_2$, the optimal structure of \mathbf{A} is in the form of $\mathbf{A} = \mathbf{U}^* \begin{bmatrix} a & 0 \\ 0 & 0 \end{bmatrix} \mathbf{U}^H$, with $a \geq 0$.

Proof: Please refer to Appendix III. \blacksquare

Note that in other cases of \mathbf{h}_1 and \mathbf{h}_2 beyond the above two, we do not have further simplified structures for \mathbf{B} , or \mathbf{A} upon that in (12). Thus, in general we need to resort to optimization techniques to obtain the 2×2 matrix \mathbf{B} for each boundary rate-pair of $\mathcal{R}(p_1, p_2, P_R)$, as will be shown next.

B. Optimization Problems

Since $\mathcal{R}(p_1, p_2, P_R)$ in (13) is the same as that in (8), we use (13) in this subsection to characterize all the boundary rate-pairs of $\mathcal{R}(p_1, p_2, P_R)$. A commonly used method to characterize different rate-tuples on the boundary of a multiuser capacity region is via solving a sequence of *weighted sum-rate maximization* (WSRMax) problems, each for a different (nonnegative) rate weight vector of users. In the case of TWRC, let $\mathbf{w} = [w_{21}, w_{12}]^T$ be the weight vector, where w_{21} and w_{12} are the “rate rewards” for r_{21} and r_{12} , respectively. From (13), we can express the WSRMax problem to determine a particular boundary rate-pair of $\mathcal{R}(p_1, p_2, P_R)$ as

$$\begin{aligned} \underset{\mathbf{B}}{\text{Max.}} \quad & \frac{w_{21}}{2} \log_2 \left(1 + \frac{|\mathbf{g}_1^T \mathbf{B} \mathbf{g}_2|^2 p_2}{\|\mathbf{B}^H \mathbf{g}_1^*\|^2 + 1} \right) \\ & + \frac{w_{12}}{2} \log_2 \left(1 + \frac{|\mathbf{g}_2^T \mathbf{B} \mathbf{g}_1|^2 p_1}{\|\mathbf{B}^H \mathbf{g}_2^*\|^2 + 1} \right) \\ \text{s.t.} \quad & \|\mathbf{B}\mathbf{g}_1\|^2 p_1 + \|\mathbf{B}\mathbf{g}_2\|^2 p_2 + \text{tr}(\mathbf{B}\mathbf{B}^H) \leq P_R. \end{aligned} \quad (14)$$

In the above problem, although the constraint is convex, the objective function is not a concave function of \mathbf{B} . As a result, this problem is non-convex [25], and is thus difficult to solve via standard convex optimization techniques.

Therefore, we need to resort to an alternative method of WSRMax to characterize $\mathcal{R}(p_1, p_2, P_R)$. In [27], an interesting concept so-called *rate profile* was introduced to efficiently characterize boundary rate-tuples of a capacity region. A rate profile regulates the ratio between each user's rate, r_k , and their sum-rate, $R_{\text{sum}} = \sum_{k=1}^K r_k$, to be a predefined value α_k , i.e., $\frac{r_k}{R_{\text{sum}}} = \alpha_k$, $k = 1, \dots, K$, with K denoting the number of users. The rate-profile vector is then defined as $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_K]^T$. For a given $\boldsymbol{\alpha}$, if R_{sum} is maximized subject to the rate-profile constraint specified by $\boldsymbol{\alpha}$, the solution rate-tuple, $R_{\text{sum}}\boldsymbol{\alpha}$, can then be geometrically viewed as the intersection of a straight line specified by a slope of $\boldsymbol{\alpha}$ and passing through the origin of the capacity region, with the capacity region boundary. Thereby, with different $\boldsymbol{\alpha}$'s, all the boundary rate-tuples of the capacity region can be obtained.

Next, we show that by applying the above method based on rate profile, boundary rate-pairs of $\mathcal{R}(p_1, p_2, P_R)$ can be efficiently characterized. Since in our case $\mathcal{R}(p_1, p_2, P_R)$ lies in a two-dimensional space, we can express the rate-profile vector as $\alpha = [\alpha_{21}, \alpha_{12}]^T$, where $\alpha_{21} = \frac{r_{21}}{R_{\text{sum}}}$, $\alpha_{12} = \frac{r_{12}}{R_{\text{sum}}}$, and $R_{\text{sum}} = r_{21} + r_{12}$. For a fixed α , we consider the following sum-rate maximization problem:

$$\begin{aligned} & \underset{\substack{R_{\text{sum}}, \mathbf{B}}} {\text{Max.}} \quad R_{\text{sum}} \\ & \text{s.t.} \quad \frac{1}{2} \log_2 \left(1 + \frac{|\mathbf{g}_1^T \mathbf{B} \mathbf{g}_2|^2 p_2}{\|\mathbf{B}^H \mathbf{g}_1^*\|^2 + 1} \right) \geq \alpha_{21} R_{\text{sum}} \\ & \quad \frac{1}{2} \log_2 \left(1 + \frac{|\mathbf{g}_2^T \mathbf{B} \mathbf{g}_1|^2 p_1}{\|\mathbf{B}^H \mathbf{g}_2^*\|^2 + 1} \right) \geq \alpha_{12} R_{\text{sum}} \\ & \quad \|\mathbf{B} \mathbf{g}_1\|^2 p_1 + \|\mathbf{B} \mathbf{g}_2\|^2 p_2 + \text{tr}(\mathbf{B} \mathbf{B}^H) \leq P_R. \end{aligned} \quad (15)$$

After solving the above problem, solution of \mathbf{B} can be used to construct the optimal relay beamforming matrix \mathbf{A} according to (12), and the solution $R_{\text{sum}} \alpha$ becomes the rate-pair on the boundary of $\mathcal{R}(p_1, p_2, P_R)$ corresponding to the given α . To solve problem (15), we first consider the following relay power minimization problem subject to rate constraints:

$$\begin{aligned} & \underset{\mathbf{B}}{\text{Min.}} \quad p_R := \|\mathbf{B} \mathbf{g}_1\|^2 p_1 + \|\mathbf{B} \mathbf{g}_2\|^2 p_2 + \text{tr}(\mathbf{B} \mathbf{B}^H) \\ & \text{s.t.} \quad \frac{1}{2} \log_2 \left(1 + \frac{|\mathbf{g}_1^T \mathbf{B} \mathbf{g}_2|^2 p_2}{\|\mathbf{B}^H \mathbf{g}_1^*\|^2 + 1} \right) \geq \alpha_{21} r \\ & \quad \frac{1}{2} \log_2 \left(1 + \frac{|\mathbf{g}_2^T \mathbf{B} \mathbf{g}_1|^2 p_1}{\|\mathbf{B}^H \mathbf{g}_2^*\|^2 + 1} \right) \geq \alpha_{12} r. \end{aligned} \quad (16)$$

If the above problem is feasible, its optimal value, denoted by p_R^* , will be the minimum relay power required to support the given rate-pair $r\alpha$; otherwise, there is no finite relay power that can support this rate-pair, and for convenience we denote $p_R^* = +\infty$ in this case. Problems (15) and (16) are related as follows. If for some given r , α and P_R , the optimal value of problem (16) satisfies that $p_R^* > P_R$, it follows that r must be an infeasible solution of R_{sum} in problem (15), i.e., the rate-pair $r\alpha$ must fall outside $\mathcal{R}(p_1, p_2, P_R)$ along the line specified by slope α ; if $p_R^* \leq P_R$, it follows that r is a feasible solution of R_{sum} and thus $r\alpha$ must be within $\mathcal{R}(p_1, p_2, P_R)$. Based on the above observations, we obtain the following algorithm for problem (15), for which a rigorous proof is given in Appendix IV.

Algorithm 3.1:

- Given $R_{\text{sum}} \in [0, \bar{R}_{\text{sum}}]$, α .
- Initialize $r_{\min} = 0$, $r_{\max} = \bar{R}_{\text{sum}}$.
- Repeat
 1. Set $r \leftarrow \frac{1}{2}(r_{\min} + r_{\max})$.
 2. Solve problem (16) to obtain its optimal value, p_R^* .
 3. Update r by the bisection method [25]: If $p_R^* \leq P_R$, set $r_{\min} \leftarrow r$; otherwise, $r_{\max} \leftarrow r$.
- Until $r_{\max} - r_{\min} \leq \delta_r$, where δ_r is a small positive constant to control the algorithm accuracy. The converged value of r_{\min} is the optimal solution of R_{sum} in (15).

Note that \bar{R}_{sum} is an upper bound on the optimal solution of R_{sum} in (15) for the given α . In Section IV (see Remark 5.1), we obtain such an upper bound that is valid for all

possible values of α . In the next subsection, we will address the remaining part in Algorithm 3.1 on how to solve problem (16) in Step 2.

C. Power Minimization under SNR Constraints

Denote γ_1 and γ_2 as the SNRs at the receivers of S1 and S2, respectively, which are defined as

$$\gamma_1 = \frac{|\mathbf{g}_1^T \mathbf{B} \mathbf{g}_2|^2 p_2}{\|\mathbf{B}^H \mathbf{g}_1^*\|^2 + 1}, \quad \gamma_2 = \frac{|\mathbf{g}_2^T \mathbf{B} \mathbf{g}_1|^2 p_1}{\|\mathbf{B}^H \mathbf{g}_2^*\|^2 + 1}. \quad (17)$$

Let $\bar{\gamma}_1 = 2^{2\alpha_{21}r} - 1$ and $\bar{\gamma}_2 = 2^{2\alpha_{12}r} - 1$ be the equivalent SNR targets at S1 and S2 to guarantee the given rate constraints. Then, it is observed that the rate constraints in (16) can be expressed as the corresponding SNR constraints at S1 and S2, $\gamma_1 \geq \bar{\gamma}_1$ and $\gamma_2 \geq \bar{\gamma}_2$, respectively. Using (17), problem (16) can be recast as the following equivalent problem:

$$\begin{aligned} & \underset{\mathbf{B}}{\text{Min.}} \quad p_R := \|\mathbf{B} \mathbf{g}_1\|^2 p_1 + \|\mathbf{B} \mathbf{g}_2\|^2 p_2 + \text{tr}(\mathbf{B} \mathbf{B}^H) \\ & \text{s.t.} \quad |\mathbf{g}_1^T \mathbf{B} \mathbf{g}_2|^2 \geq \frac{\bar{\gamma}_1}{p_2} \|\mathbf{B}^H \mathbf{g}_1^*\|^2 + \frac{\bar{\gamma}_1}{p_2} \\ & \quad |\mathbf{g}_2^T \mathbf{B} \mathbf{g}_1|^2 \geq \frac{\bar{\gamma}_2}{p_1} \|\mathbf{B}^H \mathbf{g}_2^*\|^2 + \frac{\bar{\gamma}_2}{p_1}. \end{aligned} \quad (18)$$

Note that the above problem may be of practical interest itself, since it is relevant when certain prescribed transmission quality-of-service (QoS) requirements in terms of receiver SNRs need to be fulfilled at S1 and S2. For the convenience of analysis, we modify the above problem as follows. First, let $\text{Vec}(\mathbf{Q})$ be a $K^2 \times 1$ vector associated with a $K \times K$ square matrix $\mathbf{Q} = [\mathbf{q}_1, \dots, \mathbf{q}_K]^T$, where $\mathbf{q}_k \in \mathbb{C}^{K \times 1}$, $k = 1, \dots, K$, by the rule of $\text{Vec}(\mathbf{Q}) = [\mathbf{q}_1^T, \dots, \mathbf{q}_K^T]^T$. Next, with $\mathbf{b} = \text{Vec}(\mathbf{B})$ and $\Theta = p_1 \mathbf{g}_1 \mathbf{g}_1^H + p_2 \mathbf{g}_2 \mathbf{g}_2^H + \mathbf{I}$, we can express p_R in the objective function of (18) as $p_R = \text{tr}(\mathbf{B} \Theta \mathbf{B}^H) = \|\Phi \mathbf{b}\|^2$, where $\Phi = (\text{diag}(\Theta^T, \Theta^T))^{\frac{1}{2}}$. Similarly, let $\mathbf{f}_1 = \text{Vec}(\mathbf{g}_1 \mathbf{g}_2^T)$ and $\mathbf{f}_2 = \text{Vec}(\mathbf{g}_2 \mathbf{g}_1^T)$. Then, from (18) it follows that $|\mathbf{g}_1^T \mathbf{B} \mathbf{g}_2|^2 = |\mathbf{f}_1^T \mathbf{b}|^2$ and $|\mathbf{g}_2^T \mathbf{B} \mathbf{g}_1|^2 = |\mathbf{f}_2^T \mathbf{b}|^2$. Furthermore, by defining

$$\mathbf{G}_i = \begin{bmatrix} \mathbf{g}_i(1, 1) & 0 & \mathbf{g}_i(2, 1) & 0 \\ 0 & \mathbf{g}_i(1, 1) & 0 & \mathbf{g}_i(2, 1) \end{bmatrix}, \quad i = 1, 2,$$

we have $\|\mathbf{B}^H \mathbf{g}_i^*\|^2 = \|\mathbf{G}_i \mathbf{b}\|^2$, $i = 1, 2$. Using the above transformations, (18) can be rewritten as

$$\begin{aligned} & \underset{\mathbf{b}}{\text{Min.}} \quad p_R := \|\Phi \mathbf{b}\|^2 \\ & \text{s.t.} \quad |\mathbf{f}_1^T \mathbf{b}|^2 \geq \frac{\bar{\gamma}_1}{p_2} \|\mathbf{G}_1 \mathbf{b}\|^2 + \frac{\bar{\gamma}_1}{p_2} \\ & \quad |\mathbf{f}_2^T \mathbf{b}|^2 \geq \frac{\bar{\gamma}_2}{p_1} \|\mathbf{G}_2 \mathbf{b}\|^2 + \frac{\bar{\gamma}_2}{p_1}. \end{aligned} \quad (19)$$

The above problem can be shown to be still non-convex. However, in the following, we show that the exact optimal solution could be obtained via a relaxed semidefinite programming (SDP) [25] problem.

We first define $\mathbf{E}_0 = \Phi^H \Phi$, $\mathbf{E}_1 = \frac{p_2}{\bar{\gamma}_1} \mathbf{f}_1^* \mathbf{f}_1^T - \mathbf{G}_1^H \mathbf{G}_1$, and $\mathbf{E}_2 = \frac{p_1}{\bar{\gamma}_2} \mathbf{f}_2^* \mathbf{f}_2^T - \mathbf{G}_2^H \mathbf{G}_2$. Since standard SDP formulations only involve real variables and constants, we introduce a new real matrix variable as $\mathbf{X} = [\mathbf{b}_R; \mathbf{b}_I] \times [\mathbf{b}_R; \mathbf{b}_I]^T$, where $\mathbf{b}_R = \text{Re}(\mathbf{b})$ and $\mathbf{b}_I = \text{Im}(\mathbf{b})$ are the real and imaginary parts of

\mathbf{b} , respectively. To rewrite the norm representations at (19) in terms of \mathbf{X} , we need to rewrite \mathbf{E}_0 , \mathbf{E}_1 , and \mathbf{E}_2 , as expanded matrices \mathbf{F}_0 , \mathbf{F}_1 , and \mathbf{F}_2 , respectively, in terms of their real and imaginary parts. Specifically, to write out \mathbf{F}_0 , we first define the short notations $\Phi_R = \text{Re}(\Phi)$ and $\Phi_I = \text{Im}(\Phi)$; then we have

$$\mathbf{F}_0 = \begin{bmatrix} \Phi_R^T \Phi_R + \Phi_I^T \Phi_I & \Phi_I^T \Phi_R - \Phi_R^T \Phi_I \\ \Phi_R^T \Phi_I - \Phi_I^T \Phi_R & \Phi_R^T \Phi_R + \Phi_I^T \Phi_I \end{bmatrix}.$$

The expanded matrices \mathbf{F}_1 and \mathbf{F}_2 can be generated from \mathbf{E}_1 and \mathbf{E}_2 in a similar way, where the two terms in \mathbf{E}_1 or \mathbf{E}_2 could first be expanded separately then summed together.

As such, problem (19) can be equivalently rewritten as

$$\begin{aligned} \underset{\mathbf{X}}{\text{Min.}} \quad p_R &:= \text{tr}(\mathbf{F}_0 \mathbf{X}) \\ \text{s.t.} \quad \text{tr}(\mathbf{F}_1 \mathbf{X}) &\geq 1, \quad \text{tr}(\mathbf{F}_2 \mathbf{X}) \geq 1, \quad \mathbf{X} \succeq 0, \\ \text{rank}(\mathbf{X}) &= 1. \end{aligned} \quad (20)$$

The above problem is still not convex given the last rank-one constraint. However, if we remove such a constraint, this problem is relaxed into a convex SDP problem as shown below.

$$\begin{aligned} \underset{\mathbf{X}}{\text{Min.}} \quad p_R &:= \text{tr}(\mathbf{F}_0 \mathbf{X}) \\ \text{s.t.} \quad \text{tr}(\mathbf{F}_1 \mathbf{X}) &\geq 1, \quad \text{tr}(\mathbf{F}_2 \mathbf{X}) \geq 1, \quad \mathbf{X} \succeq 0. \end{aligned} \quad (21)$$

Given the convexity of the above SDP problem, the optimal solution could be efficiently found by various convex optimization methods [25]. Note that if problem (21) is infeasible, so is the more restricted problem (20). Thus, we assume w.l.o.g. that problem (21) is feasible in the following discussions. SDP relaxation usually leads to an optimal \mathbf{X} for problem (21) that is of rank r with $r \geq 1$, which makes it impossible to reconstruct the exact optimal solution for problem (19) when $r > 1$. A commonly adopted method in the literature to obtain a feasible rank-one (but in general suboptimal) solution from the solution of SDP relaxation is via “randomization” (see, e.g., [28] and references therein). Fortunately, we show in the following that with the special structure in problem (21), we could efficiently reconstruct an optimal rank-one solution from its optimal solution that could be of rank r with $r > 1$, based on some elegant results derived for SDP relaxation in [29]. In other words, we could obtain the exact optimal solution for the non-convex problem in (20) without losing any optimality, and as efficiently as solving a convex problem.

Theorem 3.2: Assume that an optimal solution \mathbf{X}^* of rank $r > 1$ has been found for problem (21), we could efficiently construct another feasible optimal solution \mathbf{X}^{**} of rank one, i.e., \mathbf{X}^{**} is the optimal solution for both (20) and (21).

Proof: Please refer to Appendix V. ■

Since the above proof is self-constructive, we could write a routine to obtain an optimal rank-one solution for problem (20) from \mathbf{X}^* , as given in the last part of Appendix V.

IV. LOW-COMPLEXITY RELAY BEAMFORMING SCHEMES

In this section, we present suboptimal relay beamforming schemes that require lower complexity for implementation

than the optimal scheme developed in Section III. Two suboptimal beamforming structures for \mathbf{A} are proposed as follows:

- Maximal-Ratio Reception and Maximal-Ratio Transmission (MRR-MRT):

$$\mathbf{A}_{\text{MR}} = \mathbf{H}_{\text{DL}}^H \begin{bmatrix} a_{\text{MR}} & 0 \\ 0 & b_{\text{MR}} \end{bmatrix} \mathbf{H}_{\text{UL}}^H; \quad (22)$$

- Zero-Forcing Reception and Zero-Forcing Transmission (ZFR-ZFT):³

$$\mathbf{A}_{\text{ZF}} = \mathbf{H}_{\text{DL}}^\dagger \begin{bmatrix} a_{\text{ZF}} & 0 \\ 0 & b_{\text{ZF}} \end{bmatrix} \mathbf{H}_{\text{UL}}^\dagger. \quad (23)$$

Note that from (10), it follows that a_x and b_x , $x = \text{MR}$ or ZF , $a_x \geq 0$ and $b_x \geq 0$, in the above beamforming structures play the role of balancing relay power allocations to transmissions from S1 to S2 and from S2 to S1. MRR-MRT applies the “matched-filter (MF)” -based receive and transmit beamforming at R to maximize the total signal power forwarded to S1 and S2. However, in this scheme, R does not attempt to suppress or mitigate the interference between S1 and S2. On the other hand, ZFR-ZFT applies the “zero-forcing (ZF)” -based receive and transmit beamforming to remove the interferences between S1 and S2 at R as well as at the end receivers of S1 and S2. To illustrate this, we substitute \mathbf{A}_{ZF} in (23) into (10) to obtain $\begin{bmatrix} y_2(n) \\ y_1(n) \end{bmatrix}$ in the form of

$$\begin{bmatrix} a_{\text{ZF}} \sqrt{p_1} s_1(n) \\ b_{\text{ZF}} \sqrt{p_2} s_2(n) \end{bmatrix} + \begin{bmatrix} a_{\text{ZF}} & 0 \\ 0 & b_{\text{ZF}} \end{bmatrix} \mathbf{H}_{\text{UL}}^\dagger \mathbf{z}_R(n) + \begin{bmatrix} z_2(n) \\ z_1(n) \end{bmatrix}.$$

It is observed from the above that the self-interferences are completely removed at the receivers of S1 and S2 by ZF-based relay beamforming. Therefore, the main advantage of ZFR-ZFT over MRR-MRT lies in that it does not need to implement the self-interference cancelation at S1 or S2, and thus simplifies their receivers. In general, with ANC, we know that the interference between S1 and S2 observed at R is in fact the self-interference of S1 or S2, and can be later removed at the end receiver of S1 or S2. Thus, it is conjectured that MRR-MRT may have a superior performance over ZFR-ZFT for ANC-based TWRC. This conjecture is in fact true, and will be verified in later parts of this paper via performance analysis and simulation results.

Interestingly, the above two suboptimal beamforming schemes both comply with the optimal beamforming structure given in (12), while their associated values of \mathbf{B} are in general suboptimal. This can be easily verified by rewriting \mathbf{A}_{ZF} in (22) and \mathbf{A}_{MR} in (23) as $\mathbf{A}_{\text{MR}} = \mathbf{U}^* \mathbf{B}_{\text{MR}} \mathbf{U}^H$ and $\mathbf{A}_{\text{ZF}} = \mathbf{U}^* \mathbf{B}_{\text{ZF}} \mathbf{U}^H$, respectively, where

$$\mathbf{B}_{\text{MR}} = \mathbf{\Sigma} \mathbf{V}^T \begin{bmatrix} 0 & a_{\text{MR}} \\ b_{\text{MR}} & 0 \end{bmatrix} \mathbf{V} \mathbf{\Sigma} \quad (24)$$

$$\mathbf{B}_{\text{ZF}} = \mathbf{\Sigma}^{-1} \mathbf{V}^T \begin{bmatrix} 0 & a_{\text{ZF}} \\ b_{\text{ZF}} & 0 \end{bmatrix} \mathbf{V} \mathbf{\Sigma}^{-1}. \quad (25)$$

Using Lemmas 3.1 and 3.2, we can show the following results on the optimality of MRR-MRT and ZFR-ZFT in some special channel cases. For brevity, here we omit the proofs.

³Note that the ZFR-ZFT scheme with $a_{\text{ZF}} = b_{\text{ZF}}$ has also been proposed in [18], but without detailed performance analysis.

Lemma 4.1: In both cases of $\mathbf{h}_1 \perp \mathbf{h}_2$ and $\mathbf{h}_1 \parallel \mathbf{h}_2$, \mathbf{A}_{MR} in (22) is equivalent to the optimal \mathbf{A} given in (12).

Lemma 4.2: In the case of $\mathbf{h}_1 \perp \mathbf{h}_2$, \mathbf{A}_{ZF} in (23) is equivalent to the optimal \mathbf{A} given in (12).

It is also noted that in the case of $\mathbf{h}_1 \parallel \mathbf{h}_2$, \mathbf{B}_{ZF} in (25) does not exist since in Σ , $\sigma_2 = 0$, and thus Σ is non-invertible. As a result, \mathbf{A}_{ZF} does not exist either in this case.

V. PERFORMANCE ANALYSIS

To further investigate the performances of the proposed optimal and suboptimal relay beamforming schemes, we study in this section their achievable sum-rates in TWRC. First, we derive an upper bound on the maximum sum-rate or the sum-capacity achievable by the optimal beamforming scheme, as well as various lower bounds on the achievable sum-rates by the suboptimal schemes. Then, by comparing these rate bounds at asymptotically high SNR, we characterize the limiting sum-rate losses resulted by the suboptimal beamforming schemes as compared to the sum-capacity.

A. Rate Bounds

First, we study the sum-capacity of TWRC with given P_R , p_1 , and p_2 . The sum-capacity of TWRC can be obtained by solving the WSRMax problem (14) with $w_{12} = w_{21} = 1$. Since WSRMax for TWRC is non-convex and is thus difficult to solve, we consider an upper bound on the sum-capacity, which can be obtained by solving the following modified problem of (14):

$$\begin{aligned} \underset{\mathbf{B}_{21}, \mathbf{B}_{12}}{\text{Max.}} \quad & \frac{1}{2} \log_2 \left(1 + \frac{|\mathbf{g}_1^T \mathbf{B}_{21} \mathbf{g}_2|^2 p_2}{\|\mathbf{B}_{21}^H \mathbf{g}_1^*\|^2 + 1} \right) \\ & + \frac{1}{2} \log_2 \left(1 + \frac{|\mathbf{g}_2^T \mathbf{B}_{12} \mathbf{g}_1|^2 p_1}{\|\mathbf{B}_{12}^H \mathbf{g}_2^*\|^2 + 1} \right) \\ \text{s.t. } & \|\mathbf{B}_{12} \mathbf{g}_1\|^2 p_1 + \|\mathbf{B}_{21} \mathbf{g}_2\|^2 p_2 + \kappa_{12} \text{tr}(\mathbf{B}_{12} \mathbf{B}_{12}^H) \\ & + \kappa_{21} \text{tr}(\mathbf{B}_{21} \mathbf{B}_{21}^H) \leq P_R \end{aligned} \quad (26)$$

where κ_{12} and κ_{21} are nonnegative and satisfy $\kappa_{12} + \kappa_{21} = 1$. Let $C_{\text{sum}}(\kappa_{12}, \kappa_{21})$ denote the maximum value of the above problem. Note that if we add the constraint $\mathbf{B}_{12} = \mathbf{B}_{21} = \mathbf{B}$ into the above problem, solution of \mathbf{B} will lead to the exact sum-capacity of TWRC. Since $C_{\text{sum}}(\kappa_{12}, \kappa_{21})$ is an upper bound on the sum-capacity for any feasible κ_{12} and κ_{21} , it can be tightened by minimizing $C_{\text{sum}}(\kappa_{12}, \kappa_{21})$ over all the feasible pairs of κ_{12} and κ_{21} . For given κ_{12} and κ_{21} , problem (26) can be decomposed into the following two independent subproblems:

$$\begin{aligned} \underset{\mathbf{B}_{21}}{\text{Max.}} \quad & \frac{1}{2} \log_2 \left(1 + \frac{|\mathbf{g}_1^T \mathbf{B}_{21} \mathbf{g}_2|^2 p_2}{\|\mathbf{B}_{21}^H \mathbf{g}_1^*\|^2 + 1} \right) \\ \text{s.t. } & \|\mathbf{B}_{21} \mathbf{g}_2\|^2 p_2 + \kappa_{21} \text{tr}(\mathbf{B}_{21} \mathbf{B}_{21}^H) \leq P_{21} \end{aligned} \quad (27)$$

$$\begin{aligned} \underset{\mathbf{B}_{12}}{\text{Max.}} \quad & \frac{1}{2} \log_2 \left(1 + \frac{|\mathbf{g}_2^T \mathbf{B}_{12} \mathbf{g}_1|^2 p_1}{\|\mathbf{B}_{12}^H \mathbf{g}_2^*\|^2 + 1} \right) \\ \text{s.t. } & \|\mathbf{B}_{12} \mathbf{g}_1\|^2 p_1 + \kappa_{12} \text{tr}(\mathbf{B}_{12} \mathbf{B}_{12}^H) \leq P_{12} \end{aligned} \quad (28)$$

subject to an additional common constraint $P_{21} + P_{12} \leq P_R$. Let $C_{21}(\kappa_{21}, P_{21})$ and $C_{12}(\kappa_{12}, P_{12})$ denote the maximum

values of the above two subproblems, respectively. Thus, $C_{\text{sum}}(\kappa_{12}, \kappa_{21})$ can be obtained by first solving the above two subproblems for given P_{12} and P_{21} , and then maximizing $C_{21}(\kappa_{21}, P_{21}) + C_{12}(\kappa_{12}, P_{12})$ over all the feasible values of P_{12} and P_{21} . Note that each of the above two subproblems optimizes the relay beamforming matrix to maximize the capacity of the corresponding OWRC from S2 and S1 via R, or from S1 to S2 via R. By applying the results in prior work [20], [21], the optimal solutions to (27) and (28) can be obtained as

$$\mathbf{B}_{21} = \sqrt{\frac{P_{21}}{p_2 \|\mathbf{g}_2\|^2 + \kappa_{21}}} \tilde{\mathbf{g}}_1^* \tilde{\mathbf{g}}_2^H \quad (29)$$

$$\mathbf{B}_{12} = \sqrt{\frac{P_{12}}{p_1 \|\mathbf{g}_1\|^2 + \kappa_{12}}} \tilde{\mathbf{g}}_2^* \tilde{\mathbf{g}}_1^H \quad (30)$$

where $\tilde{\mathbf{g}}_i = \frac{\mathbf{g}_i}{\|\mathbf{g}_i\|}, i = 1, 2$. By substituting the above expressions into the objective functions of (27) and (28), respectively, we obtain

$$C_{21}(\kappa_{21}, P_{21}) = \frac{1}{2} \log_2 \left(1 + \frac{\theta_2 p_2}{1 + \frac{(\theta_2/\theta_1)p_2}{P_{21}} + \frac{\kappa_{21}}{\theta_1 P_{21}}} \right) \quad (31)$$

$$C_{12}(\kappa_{12}, P_{12}) = \frac{1}{2} \log_2 \left(1 + \frac{\theta_1 p_1}{1 + \frac{(\theta_1/\theta_2)p_1}{P_{12}} + \frac{\kappa_{12}}{\theta_2 P_{12}}} \right) \quad (32)$$

where, for conciseness, we have denoted $\|\mathbf{g}_1\|^2 = \|\mathbf{h}_1\|^2 = \theta_1$ and $\|\mathbf{g}_2\|^2 = \|\mathbf{h}_2\|^2 = \theta_2$. It then follows that the tightest upper bound on the sum-capacity, denoted as C_{UB} , can be obtained as

$$C_{\text{UB}} = \min_{\kappa_{21} + \kappa_{12} = 1} \max_{P_{21} + P_{12} \leq P_R} C_{21}(\kappa_{21}, P_{21}) + C_{12}(\kappa_{12}, P_{12}). \quad (33)$$

Unfortunately, there is in general no closed-form solution of C_{UB} , and thus numerical search over all the feasible values of $\kappa_{21}, \kappa_{12}, P_{21}$, and P_{12} is needed to obtain C_{UB} . Since $C_{21}(\kappa_{21}, P_{21})$ and $C_{12}(\kappa_{12}, P_{12})$ are increasing functions of P_{21} and P_{12} , respectively, a simple upper bound on the sum-capacity (less tighter than C_{UB}) can be obtained from (31) and (32) with $\kappa_{21} = \kappa_{12} = 1/2$ and $P_{21} = P_{12} = P_R$ as follows:

$$\begin{aligned} C_{\text{UB}}^{(0)} = & \frac{1}{2} \log_2 \left(1 + \frac{\theta_2 p_2}{1 + \frac{(\theta_2/\theta_1)p_2}{P_R} + \frac{1}{2\theta_1 P_R}} \right) \\ & + \frac{1}{2} \log_2 \left(1 + \frac{\theta_1 p_1}{1 + \frac{(\theta_1/\theta_2)p_1}{P_R} + \frac{1}{2\theta_2 P_R}} \right). \end{aligned} \quad (34)$$

Remark 5.1: Note that $C_{\text{UB}}^{(0)}$ given in (34) can be used as \bar{R}_{sum} for Algorithm 3.1 in Section III. Since $C_{\text{UB}}^{(0)}$ is obtained without any constraint on rate allocations among r_{21} and r_{12} , it is a valid upper bound on the achievable sum-rate regardless of the rate-profile vector α .

Next, we derive the lower bounds on the sum-rates achievable by the proposed suboptimal relay beamforming schemes, MRR-MRT and ZFR-ZFT, denoted as $R_{\text{LB}}^{\text{MR}}$ and $R_{\text{LB}}^{\text{ZF}}$, respectively. Since the rate lower bound is of interest, we assume here $a_x = b_x$, where $x = \text{MR}$ in (22) or ZF in (23). For conciseness, define $\rho = \frac{|\mathbf{h}_1^H \mathbf{h}_2|^2}{\theta_1 \theta_2}$ as the correlation coefficient

between \mathbf{h}_1 and \mathbf{h}_2 . Then, the following lemmas are obtained.

Lemma 5.1: With the MRR-MRT relay beamforming scheme, the achievable sum-rate of TWRC is lower-bounded by R_{LB}^{MR} given in (35) (see next page). ■

Proof: Please refer to Appendix VI. ■

Lemma 5.2: With the ZFR-ZFT relay beamforming scheme, the achievable sum-rate of TWRC is lower-bounded by R_{LB}^{ZF} given in (36) (see next page). ■

Proof: Please refer to Appendix VII. ■

B. Asymptotic Results

Since the main advantage of TWRC over OWRC is to recover the loss of spectral efficiency due to half-duplex transmissions (see, e.g., [10]-[13]), it is important to examine the achievable sum-rate in TWRC at asymptotically high SNR. In the following theorem, asymptotic results on various upper and lower rate bounds in (34), (35), and (36) are presented.

Theorem 5.1: Let p_1 , p_2 , and P_R all go to infinity with fixed $\frac{P_R}{p_1} = K_1$ and $\frac{P_R}{p_2} = K_2$. Then, $C_{UB}^{(0)}$, R_{LB}^{MR} , and R_{LB}^{ZF} converge to the values given in (37), (38), and (39), respectively (see next page).

It is observed from Theorem 5.1 that at high SNR both MRR-MRT and ZFR-ZFT (provided that $\rho < 1$) asymptotically achieve the same sum-rate pre-log factor (sum-rate normalized by $\log_2 P_R$ as $P_R \rightarrow \infty$) as that of the sum-capacity upper bound, $C_{UB}^{(0)}$. However, they may have different rate gaps from $C_{UB}^{(0)}$, which are constants independent of P_R . In order to gain more insights on the limiting sum-rate losses of suboptimal beamforming schemes, in the following corollary, we compare the difference between $C_{UB}^{(0)}$ and R_{LB}^{MR} or R_{LB}^{ZF} at asymptotically high SNR in a “symmetric” TWRC with equal channel gains, i.e., $\|\mathbf{h}_1\|^2 = \|\mathbf{h}_2\|^2 = \theta$, and equal source and relay transmit powers, i.e., $p_1 = p_2 = P_R$. In this case, we can obtain a tighter upper bound on the sum-capacity than $C_{UB}^{(0)}$ as follows. For the symmetric TWRC, with $\kappa_{12} = \kappa_{21} = 1/2$, it can be easily verified that the maximization over P_{12} and P_{21} in (33) is achieved when $P_{12} = P_{21} = P_R/2$ and as a result a tighter upper bound over $C_{UB}^{(0)}$ for the symmetric TWRC is obtained as

$$C_{UB}^{(S)} = \log_2 \left(1 + \frac{\theta P_R}{3 + \frac{1}{\theta P_R}} \right). \quad (40)$$

Corollary 5.1: At asymptotically high SNR, under the assumptions that $\theta_1 = \theta_2$ and $K_1 = K_2 = 1$, we have

$$C_{UB}^{(S)} - R_{LB}^{MR} = \log_2 \left(\frac{1 + 3\rho}{(1 + \rho)^2} \right) \quad (41)$$

$$C_{UB}^{(S)} - R_{LB}^{ZF} = \log_2 \left(\frac{1}{1 - \rho} \right). \quad (42)$$

It is noted that for $0 \leq \rho \leq 1$, $\frac{1+3\rho}{(1+\rho)^2}$ has the minimum value equal to 0 at $\rho = 0$ or 1, and the maximum value equal to $9/8$ at $\rho = 1/3$. Therefore, from (41), it follows that the sum-rate loss of MRR-MRT from the sum-capacity is at most $\log_2(9/8) \approx 0.1699$ bits/complex dimension at asymptotically high SNR. On the other hand, it is observed from (42) that the sum-rate loss resulted by ZFR-ZFT increases with ρ , or when

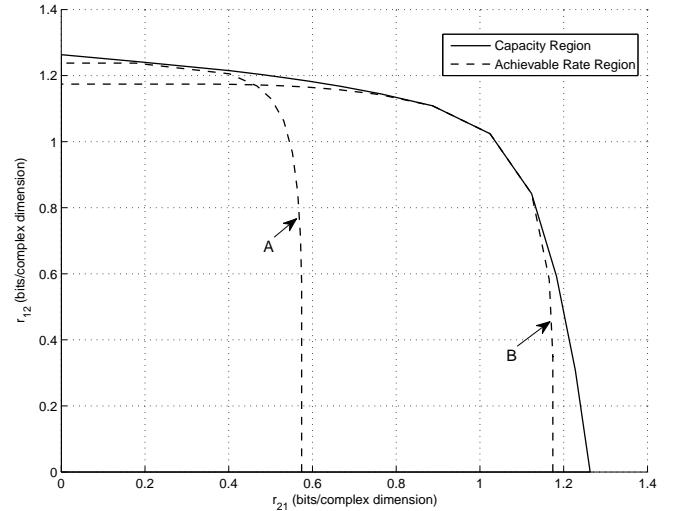


Fig. 2. Capacity region of the ANC-based TWRC with $M = 4$, $P_1 = P_2 = P_R = 10$, and $\rho = 0.5$. Note that the two rate regions enclosed by the dashed lines are example achievable rate regions $\mathcal{R}(p_1, p_2, P_R)$'s defined in (8), each with some fixed p_1 and p_2 , $p_1 \leq P_1$ and $p_2 \leq P_2$. The achievable rate region denoted by A corresponds to $p_1 = P_1$ and $p_2 < P_2$, while that denoted by B corresponds to $p_1 = P_1$ and $p_2 = P_2$.

\mathbf{h}_1 and \mathbf{h}_2 become more correlated. This is intuitively correct, since with the increasing channel correlation, more SNR loss will be incurred to separate the signals from/to S1 and S2 at R by ZF-based receive/transmit beamforming. Also note that for MRR-MRT, at $\rho = 0$ or 1, the sum-rate loss is zero, which is consistent with Lemma 4.1, while the sum-rate loss is zero for ZFR-ZFT at $\rho = 0$, which is consistent with Lemma 4.2.

VI. NUMERICAL RESULTS

In this section, we present numerical results on the achievable rates of various beamforming schemes considered in this paper, and compare them with those of other existing schemes in the literature. For convenience, we assume that \mathbf{h}_1 is a randomly generated CSCG vector $\sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$, and \mathbf{h}_1 is normalized by its own vector norm such that $\|\mathbf{h}_1\| = 1$. We then generate \mathbf{h}_2 according to $\mathbf{h}_2 = \sqrt{\rho}\mathbf{h}_1 + \sqrt{1-\rho}\mathbf{h}_w$, where \mathbf{h}_w is also a normalized CSCG random vector, $\|\mathbf{h}_w\| = 1$ and $\mathbf{h}_1^H \mathbf{h}_w = 0$. Thereby, it can be easily verified that $\|\mathbf{h}_2\| = 1$ and $\|\mathbf{h}_1^H \mathbf{h}_2\|^2 = \rho$. It is assumed that $M = 4$ in this section.

A. Capacity Region of ANC-Based TWRC

Fig. 2 shows the capacity region, $\mathcal{C}(P_1, P_2, P_R)$ defined in (9), for the ANC-based TWRC with $P_1 = P_2 = P_R = 10$, and $\rho = 0.5$. It is observed that $\mathcal{C}(P_1, P_2, P_R)$ is symmetric over r_{12} and r_{21} in this case. Notice that boundary rate-pairs of $\mathcal{C}(P_1, P_2, P_R)$ are resulted by the union over those of achievable rate regions, $\mathcal{R}(p_1, p_2, P_R)$'s defined in (8), with different values of p_1 and p_2 , $0 \leq p_1 \leq P_1$ and $0 \leq p_2 \leq P_2$. Boundary rate-pairs of each constituting $\mathcal{R}(p_1, p_2, P_R)$ are obtained by solving problem (15) using Algorithm 3.1 with different rate-profile vectors α 's. It is observed that when $p_1 = 0$ and $p_2 = P_2$, $\mathcal{R}(p_1, p_2, P_R)$ collapses into the horizontal rate axis of r_{21} , and the maximum value of r_{21} becomes the capacity of the OWRC with S2 transmitting to

$$R_{LB}^{MR} = \frac{1}{2} \log_2 \left(1 + \frac{\theta_2 p_2}{\left(1 + \frac{p_1 + (\theta_2/\theta_1)p_2}{P_R} \right) \frac{1+3\rho}{(1+\rho)^2} + \frac{2}{\theta_1(1+\rho)P_R}} \right) + \frac{1}{2} \log_2 \left(1 + \frac{\theta_1 p_1}{\left(1 + \frac{(\theta_1/\theta_2)p_1+p_2}{P_R} \right) \frac{1+3\rho}{(1+\rho)^2} + \frac{2}{\theta_2(1+\rho)P_R}} \right). \quad (35)$$

$$R_{LB}^{ZF} = \log_2 \left(1 + \frac{2p_1 p_2}{\frac{\theta_1 + \theta_2}{\theta_1 \theta_2 (1-\rho)} \left(1 + \frac{p_1}{P_R} + \frac{p_2}{P_R} \right) \left(\max(p_1, p_2) + \frac{\theta_1 + \theta_2}{\theta_1 \theta_2 (1-\rho)} \frac{p_1 + p_2}{P_R + p_1 + p_2} \right)} \right). \quad (36)$$

$$C_{UB}^{(0)} = \log_2(P_R) + \frac{1}{2} \log_2 \left(\frac{\theta_1 \theta_2}{(K_2 + \theta_2/\theta_1)(K_1 + \theta_1/\theta_2)} \right) + o(1) \quad (37)$$

$$R_{LB}^{MR} = \log_2(P_R) + \frac{1}{2} \log_2 \left(\frac{\theta_1 \theta_2}{(K_2 + K_1/K_2 + \theta_2/\theta_1)(K_1 + K_2/K_1 + \theta_1/\theta_2) \frac{(1+3\rho)^2}{(1+\rho)^4}} \right) + o(1) \quad (38)$$

$$R_{LB}^{ZF} = \log_2(P_R) + \log_2 \left(\frac{\theta_1 \theta_2}{(1 + \max(K_1, K_2) + \max(K_1/K_2, K_2/K_1)) \frac{\theta_1 + \theta_2}{2(1-\rho)}} \right) + o(1). \quad (39)$$

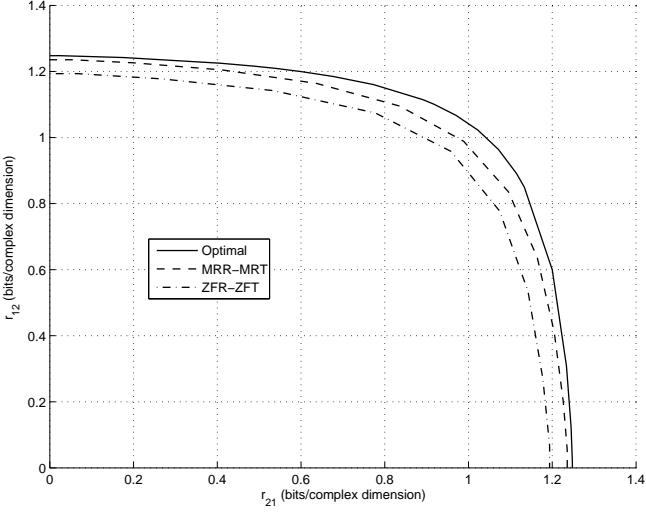


Fig. 3. Achievable rate regions of the ANC-based TWRC with $M = 4$, $p_1 = p_2 = 10$, $P_R = 10$, and $\rho = 0.1$.

S1 via R. Similarly, $\mathcal{R}(P_1, 0, P_R)$ collapses into the vertical rate axis of r_{12} , and the maximum value of r_{12} becomes the capacity of the OWRC with S1 transmitting to S2 via R.

B. Achievable Rates of Suboptimal Beamforming Schemes

Next, we examine the achievable rates of the proposed suboptimal relay beamforming schemes. Figs. 3, 4, and 5 show the achievable rate region, $\mathcal{R}(p_1, p_2, P_R)$, for the TWRC with different values of ρ , $\rho = 0.1, 0.5$, and 0.8 , respectively. It is assumed that transmit powers at S1 and S2 are fixed as $p_1 = p_2 = 10$, and the relay transmit power constraint is $P_R = 10$. Three relay beamforming schemes are compared in each of these figures, which are the optimal scheme (Algorithm 3.1), the MRR-MRT scheme (22), and the ZFR-ZFT scheme (23). Note that boundary rate-pairs of $\mathcal{R}(p_1, p_2, P_R)$ corresponding to MRR-MRT are obtained by changing different ratios between a_{MR} and b_{MR} in (22). $\mathcal{R}(p_1, p_2, P_R)$ for ZFR-ZFT is obtained in a similar way. It is observed that the achievable rate region by MRR-MRT is very close to that with the optimal

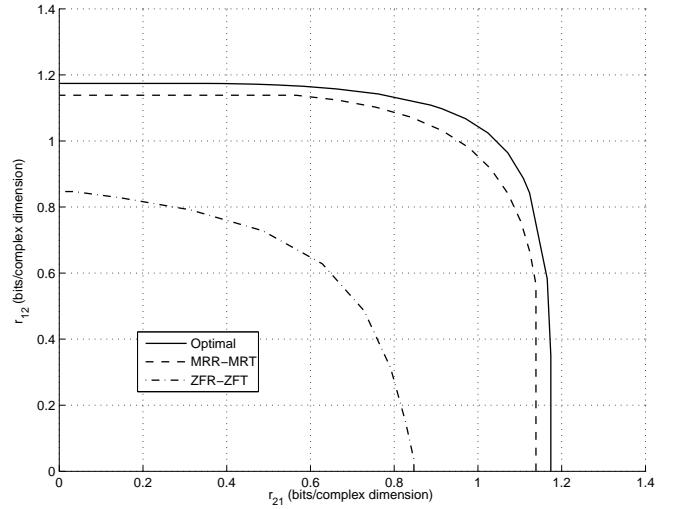


Fig. 4. Achievable rate region of the ANC-based TWRC with $M = 4$, $p_1 = p_2 = 10$, $P_R = 10$, and $\rho = 0.5$.

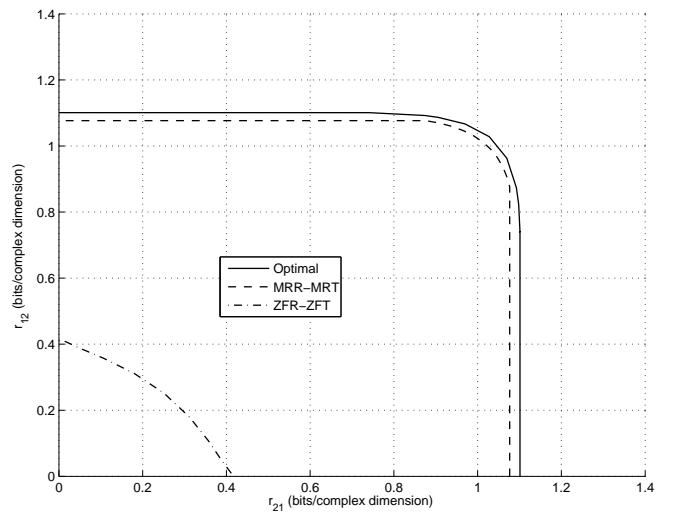


Fig. 5. Achievable rate regions of the ANC-based TWRC with $M = 4$, $p_1 = p_2 = 10$, $P_R = 10$, and $\rho = 0.8$.

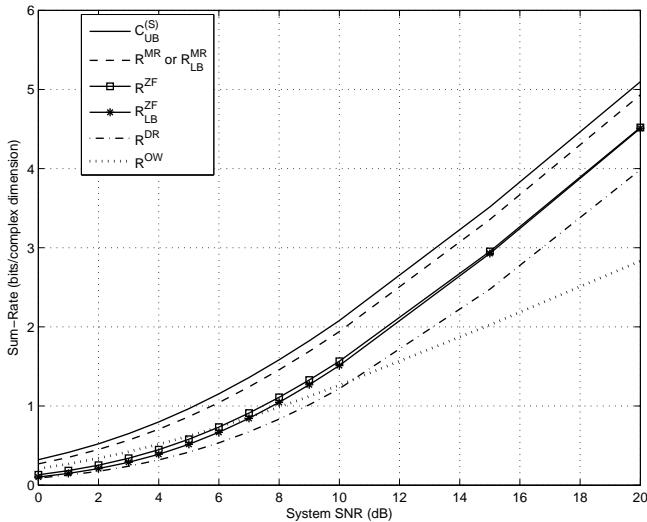


Fig. 6. Sum-rate versus system SNR for the ANC-based TWRC with $M = 4$ and $\rho = 1/3$.

scheme when the channel correlation coefficient ρ is either small or large, which is in accord with Lemma 4.1. Even for moderate values of ρ , e.g., $\rho = 0.5$, the rate loss of MRR-MRT is observed to be negligible, suggesting that MRR-MRT in fact performs very close to the optimal scheme under different channel conditions. In contrast, ZFR-ZFT performs close to the optimal scheme when ρ is small, which is in accord with Lemma 4.2. This is due to the fact that when \mathbf{h}_1 and \mathbf{h}_2 are sufficiently decorrelated, ZF-based receive/transmit beamforming at R is able to suppress the UL/DL interference between S1 and S2 with small SNR losses. However, as ρ increases, it is observed that the achievable rates of ZFR-ZFT degrade significantly as compared to those of the optimal scheme or MRR-MRT.

In Fig. 6, we show the achievable sum-rate of TWRC with $\rho = 1/3$ versus the “system” SNR. Under the assumption that transmit powers at S1, S2, and R are all equal, i.e., $p_1 = p_2 = P_R$, due to the unit-norm channels and unit-variance noises, the system SNR is conveniently set equal to P_R . Various sum-rate bounds presented in this paper are shown, including $C_{\text{UB}}^{(S)}$ in (40), $R_{\text{LB}}^{\text{MR}}$ in (35), and $R_{\text{LB}}^{\text{ZF}}$ in (36). In addition, the actual achievable sum-rates of MRR-MRT and ZFR-ZFT, denoted as R^{MR} and R^{ZF} , respectively, are also shown for comparison. Note that due to the channel symmetry, a_{MR} and b_{MR} in (22) should be equal to maximize R^{MR} ; thus, from the derivations in Appendix VI it follows that $R^{\text{MR}} = R_{\text{LB}}^{\text{MR}}$ for MRR-MRT. On the other hand, for ZFR-ZFT, a_{ZF} and b_{ZF} in (23) should also be equal to maximize R^{ZF} in this symmetric-channel case; however, from Appendix VII it follows that even with $a_{\text{ZF}} = b_{\text{ZF}}$, $R_{\text{LB}}^{\text{ZF}} < R^{\text{ZF}}$ in general, where R^{ZF} can be obtained from the RHS of (59). We also show the sum-rates of the following two heuristic schemes: (1) *Direct relaying*, where the relay beamforming matrix is in the form of $\mathbf{A} = \zeta \mathbf{I}$, with ζ being a constant determined by P_R ; (2) *One-way alternative relaying*, where four time-slots are used for one round of information exchange between S1 and S2, with two for S2 transmitting to S1 via R, and the other two for S1 to S2 via R, and the corresponding optimal relay beamforming matrices are in the

form of $\mathbf{A}_{21} = \psi \mathbf{h}_1^* \mathbf{h}_2^H$ and $\mathbf{A}_{12} = \psi \mathbf{h}_2^* \mathbf{h}_1^H$, respectively, with ψ determined by P_R [20], [21]. We denote R^{DR} and R^{OW} as the achievable sum-rates of these two schemes, respectively.

It is observed in Fig. 6 that at asymptotically high SNR, the sum-rate of MRR-MRT converges to the sum-capacity upper bound with a constant gap of 0.1699 bits/complex dimension, while ZFR-ZFT has a sum-rate gap of $\log_2(1/(1 - 1/3)) = 0.5850$ bits/complex dimension. The above observations agree with Corollary 5.1. It is also observed that the lower bound on the sum-rate by ZFR-ZFT, $R_{\text{LB}}^{\text{ZF}}$, is very tight at all SNR values. Notice that R^{DR} and R^{OW} both have significant gaps from R^{MR} at asymptotically high SNR, since the former has no beamforming gain at R, and the latter roughly incurs a loss of half the spectral efficiency due to alternative relaying.

C. Comparison with DF-Based TWRC

At last, we compare the capacity region of ANC/AF-based TWRC derived in this paper with that of DF-based TWRC recently reported in [17], for the same physical TWRC. In order to differentiate the above two capacity regions, we denote the former as \mathcal{C}_{AF} and the latter as \mathcal{C}_{DF} . Note that with DF relay operation, R first decodes both messages from S1 and S2 as in the conventional Gaussian multiple-access channel (MAC) during the first time-slot; R then re-encodes the decoded messages jointly into a new message, and transmits it over the broadcast channel (BC) to both S1 and S2 during the second time-slot. Each of S1 and S2 decodes the message of the other from the received signal given the side information on its own previously transmitted message (in the first time-slot). The achievable rates of S2 and S1 during the first MAC phase can be expressed as [30]

$$\begin{aligned} \mathcal{C}_{\text{DF}}^{\text{MAC}}(P_1, P_2) \triangleq & \left\{ (r_{21}, r_{12}) : r_{21} \leq \log_2 \left| \mathbf{I} + P_2 \mathbf{h}_2 \mathbf{h}_2^H \right|, \right. \\ & r_{12} \leq \log_2 \left| \mathbf{I} + P_1 \mathbf{h}_1 \mathbf{h}_1^H \right|, \\ & \left. r_{21} + r_{12} \leq \log_2 \left| \mathbf{I} + P_1 \mathbf{h}_1 \mathbf{h}_1^H + P_2 \mathbf{h}_2 \mathbf{h}_2^H \right| \right\}. \end{aligned} \quad (43)$$

The maximum achievable rate-pairs during the second BC phase can be expressed as [17]

$$\begin{aligned} \mathcal{C}_{\text{DF}}^{\text{BC}}(P_R) \triangleq & \bigcup_{\mathbf{S}_R: \mathbf{S}_R \succeq 0, \text{tr}(\mathbf{S}_R) \leq P_R} \left\{ (r_{21}, r_{12}) : \right. \\ & r_{21} \leq \log_2 \left(1 + \mathbf{h}_1^T \mathbf{S}_R \mathbf{h}_1^* \right), r_{12} \leq \log_2 \left(1 + \mathbf{h}_2^T \mathbf{S}_R \mathbf{h}_2^* \right) \left. \right\} \end{aligned} \quad (44)$$

where \mathbf{S}_R is the transmit signal covariance matrix at R. Note that in order to obtain $\mathcal{C}_{\text{DF}}^{\text{BC}}(P_R)$ in (44), we need to solve a sequence of optimization (WSRMax) problems expressed below with different nonnegative rate weights w_{21} and w_{12} .

$$\begin{aligned} \text{Max. } & w_{21} \log_2 \left(1 + \mathbf{h}_1^T \mathbf{S}_R \mathbf{h}_1^* \right) + w_{12} \log_2 \left(1 + \mathbf{h}_2^T \mathbf{S}_R \mathbf{h}_2^* \right) \\ \text{s.t. } & \text{tr}(\mathbf{S}_R) \leq P_R, \mathbf{S}_R \succeq 0. \end{aligned} \quad (45)$$

Since the above problem is convex, it can be solved by standard convex optimization techniques, e.g., the interior-point method [25]. Unlike the AF relay operation, DF relay

operation allows different time allocations between the MAC and BC time-slots. Let τ and $1 - \tau$ denote the percentages of the total time allocated to MAC phase and BC phase, respectively. Then, combining both MAC and BC phases yields the capacity region for DF-based TWRC, $\mathcal{C}_{\text{DF}}(P_1, P_2, P_R)$, expressed as

$$\bigcup_{\tau:0 \leq \tau \leq 1} \left(\tau \cdot \mathcal{C}_{\text{DF}}^{\text{MAC}}(P_1, P_2) \cap (1 - \tau) \cdot \mathcal{C}_{\text{DF}}^{\text{BC}}(P_R) \right). \quad (46)$$

In Figs. 7 and 8, we show $\frac{1}{2}\mathcal{C}_{\text{DF}}^{\text{MAC}}$, $\frac{1}{2}\mathcal{C}_{\text{DF}}^{\text{BC}}$, \mathcal{C}_{DF} , and \mathcal{C}_{AF} for $\rho = 0.95$ and 0.8 , respectively. It is assumed that $P_1 = P_2 = P_R = 100$. Note that in each figure, \mathcal{C}_{DF} can be visualized as the union of rate regions, each of which corresponds to the intersection of $\frac{1}{2}\mathcal{C}_{\text{DF}}^{\text{MAC}}$ and $\frac{1}{2}\mathcal{C}_{\text{DF}}^{\text{BC}}$ after they are properly scaled by 2τ and $2(1 - \tau)$, respectively, for a particular value of τ . It is observed that the DF-based TWRC in general has a larger capacity region over the AF-based counterpart. Furthermore, it is observed that this capacity gain enlarges as ρ decreases, i.e., the channels \mathbf{h}_1 and \mathbf{h}_2 become more weakly correlated. This is mainly due to the fact that when the UL channels become less correlated, R is more capable of decoding the messages from S1 and S2 during the MAC phase and as a result, $\frac{1}{2}\mathcal{C}_{\text{DF}}^{\text{MAC}}$ is observed to get enlarged as ρ decreases. If we want to draw a more fair comparison between AF- and DF-based TWRCs with the same energy consumption, we may assume that for the DF-based TWRC, equal-duration time-slots are assigned to the MAC and BC phases, i.e., $\tau = 1/2$, the same as the AF case. As such, since for both $\rho = 0.95$ and 0.8 , $\frac{1}{2}\mathcal{C}_{\text{DF}}^{\text{MAC}}$ appears as a subset of $\frac{1}{2}\mathcal{C}_{\text{DF}}^{\text{BC}}$, it concludes that the capacity region for DF-based TWRC with the fixed $\tau = 1/2$ is simply $\frac{1}{2}\mathcal{C}_{\text{DF}}^{\text{MAC}}$. Interestingly, it is observed that \mathcal{C}_{AF} improves over $\frac{1}{2}\mathcal{C}_{\text{DF}}^{\text{MAC}}$ when $\rho = 0.95$ in the region where the values of r_{21} and r_{12} are close to each other. Notice that in this region the sum-capacity in the AF case is achieved. Since DF relaying incurs larger complexity for encoding/decoding at R as compared with AF relaying, AF relaying may be a more suitable solution in practice where strong channel correlation is encountered.⁴ However, in the case of $\rho = 0.8$, it is observed that the capacity improvement of \mathcal{C}_{AF} over $\frac{1}{2}\mathcal{C}_{\text{DF}}^{\text{MAC}}$ diminishes.

VII. CONCLUSION AND FUTURE WORK

This paper studied the fundamental capacity limits of ANC/AF-based TWRC with multi-antennas at the relay. It was shown that the standard method to characterize the capacity region via WSRMax is not directly applicable to ANC-based TWRC due to the non-convexity of the optimization problem. Therefore, we proposed an alternative method to characterize the capacity region of TWRC by applying the idea of rate profile. As a byproduct, we also provided the solution for the relay power minimization problem under given SNR constraints at the receivers. Due to the bidirectional transmission as well as the self-interference cancellation by ANC, we found that the design of relay beamforming in TWRC differs very much from the conventional designs for the OWRC or the UL/DL beamforming in the traditional cellular network. We

⁴Note that in the extreme case of $\rho = 1$, the multi-antenna TWRC becomes equivalent to the single-antenna TWRC studied in [3].

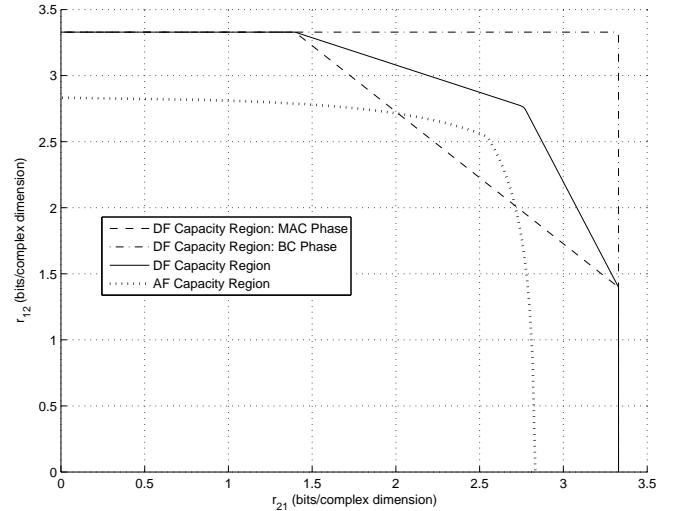


Fig. 7. Comparison of capacity region for ANC/AF-based versus DF-based TWRC with $M = 4$, $P_1 = P_2 = P_R = 100$, and $\rho = 0.95$.

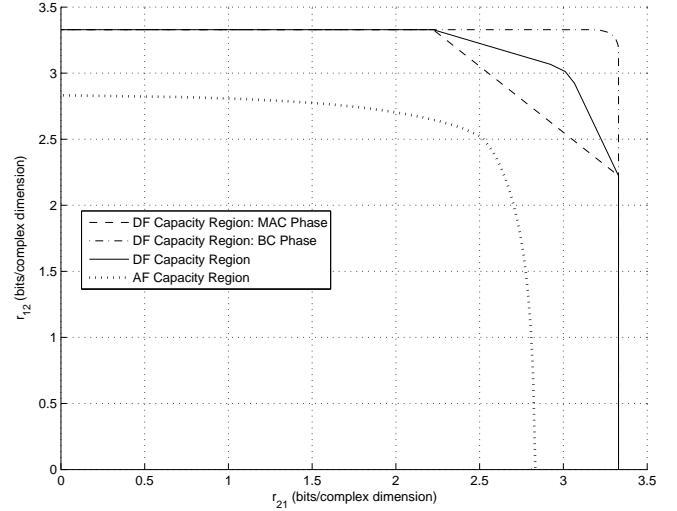


Fig. 8. Comparison of capacity region for ANC/AF-based versus DF-based TWRC with $M = 4$, $P_1 = P_2 = P_R = 100$, and $\rho = 0.8$.

presented the general form of the optimal relay beamforming structure in TWRC, as well as two low-complexity suboptimal schemes, namely, MRR-MRT and ZFR-ZFT. It was shown that ZFR-ZFT with the objective of suppressing the UL and DL interferences between S1 and S2 may not perform well in the case of strong channel correlation, while MRR-MRT with the objective of maximizing the total forwarded signal power from R to S1 and S2 achieves sum-rates and rate regions close to the optimal ones under various SNR and channel conditions. This suggests that MRR-MRT can be a good solution from an implementation viewpoint. It was also shown that the ANC/AF-based TWRC can have a capacity gain over the DF-based TWRC for sufficiently large channel correlations and equal MAC and BC time-durations.

Future work beyond this paper may include the joint design of source and relay beamforming when each source is also equipped with multi-antennas, the relay beamforming design for more than one source-pairs with different combined unicast/multicast transmissions, and the design of a hybrid AF/DF

scheme that probably improves the performances of both AF- and DF-based TWRCs. In addition, the study of estimate-and-forward (EF) relay operations for the multi-antenna TWRC is also appealing.

APPENDIX I PROOF OF THEOREM 3.1

Without loss of generality, we can express \mathbf{A} as

$$\mathbf{A} = [\mathbf{U}^*, (\mathbf{U}^\perp)^*] \begin{bmatrix} \mathbf{B} & \mathbf{C} \\ \mathbf{D} & \mathbf{E} \end{bmatrix} [\mathbf{U}, \mathbf{U}^\perp]^H \quad (47)$$

$$\begin{aligned} &= \mathbf{U}^* \mathbf{B} \mathbf{U}^H + \mathbf{U}^* \mathbf{C} (\mathbf{U}^\perp)^H + (\mathbf{U}^\perp)^* \mathbf{D} \mathbf{U}^H \\ &\quad + (\mathbf{U}^\perp)^* \mathbf{E} (\mathbf{U}^\perp)^H \end{aligned} \quad (48)$$

where $\mathbf{U}^\perp \in \mathbb{C}^{M \times (M-2)}$, $\mathbf{U}^\perp (\mathbf{U}^\perp)^H = \mathbf{I} - \mathbf{U} \mathbf{U}^H$, and \mathbf{B} , \mathbf{C} , \mathbf{D} , and \mathbf{E} are complex matrices of size 2×2 , $2 \times (M-2)$, $(M-2) \times 2$, and $(M-2) \times (M-2)$, respectively. First, it can be shown that in (6), $|\mathbf{h}_1^T \mathbf{A} \mathbf{h}_2|^2 = |\mathbf{h}_1^T \mathbf{U}^* \mathbf{B} \mathbf{U}^H \mathbf{h}_2|^2$, and $\|\mathbf{A}^H \mathbf{h}_1^*\|^2 = \|\mathbf{B}^H \mathbf{U}^T \mathbf{h}_1^*\|^2 + \|\mathbf{C}^H \mathbf{U}^T \mathbf{h}_1^*\|^2 \geq \|\mathbf{B}^H \mathbf{U}^T \mathbf{h}_1^*\|^2$. Thus, it follows that r_{21} does not depend on \mathbf{D} and \mathbf{E} , and is maximized when $\mathbf{C} = \mathbf{0}$. Similarly, from (7), we can show that r_{12} is also not related to \mathbf{D} and \mathbf{E} , and is maximized when $\mathbf{C} = \mathbf{0}$. Next, for the relay power constraint (8), from (3) it can be shown that p_R is minimized when \mathbf{C} , \mathbf{D} , and \mathbf{E} are all equal to $\mathbf{0}$. Since each rate-pair on the boundary of $\mathcal{R}(p_1, p_2, P_R)$ defined in (8) must maximize r_{21} and r_{12} subject to the given P_R , it concludes that all \mathbf{C} , \mathbf{D} , and \mathbf{E} in the corresponding \mathbf{A} should be $\mathbf{0}$. Thus, from (48), we conclude that $\mathbf{A} = \mathbf{U}^* \mathbf{B} \mathbf{U}^H$.

APPENDIX II PROOF OF LEMMA 3.1

In the case of $\mathbf{h}_1 \perp \mathbf{h}_2$, it can be easily shown that $\mathbf{g}_1 = [\|\mathbf{h}_1\|, 0]^T$ and $\mathbf{g}_2 = [0, \|\mathbf{h}_2\|]^T$. Let $\mathbf{B} = \begin{bmatrix} a & c \\ d & b \end{bmatrix}$. Substituting \mathbf{g}_1 and \mathbf{g}_2 into (13) yields

$$r_{21} \leq \frac{1}{2} \log_2 \left(1 + \frac{\|\mathbf{h}_1\|^2 \|\mathbf{h}_2\|^2 |c|^2 p_2}{\|\mathbf{h}_1\|^2 (|a|^2 + |c|^2) + 1} \right) \quad (49)$$

$$r_{12} \leq \frac{1}{2} \log_2 \left(1 + \frac{\|\mathbf{h}_1\|^2 \|\mathbf{h}_2\|^2 |d|^2 p_1}{\|\mathbf{h}_2\|^2 (|b|^2 + |d|^2) + 1} \right) \quad (50)$$

$$\begin{aligned} &\|\mathbf{h}_1\|^2 (|a|^2 + |d|^2) + \|\mathbf{h}_2\|^2 (|c|^2 + |b|^2) \\ &\quad + |a|^2 + |b|^2 + |c|^2 + |d|^2 \leq P_R. \end{aligned} \quad (51)$$

It then follows that r_{21} and r_{12} are maximized along with the relay transmit power being minimized when $|a| = 0$ and $|b| = 0$ and, thus, $a = b = 0$. Since in the above rate and power expressions only $|c|^2$ and $|d|^2$ are involved, we can assume w.l.o.g. that $c \geq 0$ and $d \geq 0$.

APPENDIX III PROOF OF LEMMA 3.2

In the case of $\mathbf{h}_1 \parallel \mathbf{h}_2$, it can be easily shown that $\mathbf{g}_1 = [\|\mathbf{h}_1\|, 0]^T$ and $\mathbf{g}_2 = [\|\mathbf{h}_2\|, 0]^T$. Let $\mathbf{B} = \begin{bmatrix} a & c \\ d & b \end{bmatrix}$. Similarly like the proof of Lemma 3.1 in Appendix II, by substituting \mathbf{g}_1 and \mathbf{g}_2 into (13), it follows that r_{21} and r_{12} are maximized along with the relay transmit power being minimized when $b = c = d = 0$, and we can assume w.l.o.g. that $a \geq 0$.

APPENDIX IV PROOF OF CONVERGENCE OF ALGORITHM 3.1

In this appendix, we prove that Algorithm 3.1 guarantees the convergence of r_{\min} to the optimal solution of problem (15). First, we show that r_{\min} is a feasible solution of problem (15): Given $R_{\text{sum}} = r_{\min}$, from Algorithm 3.1 it is easily verified that all the three constraints of problem (15) are satisfied. Secondly, suppose that there exists another feasible solution \hat{r} for problem (15) such that $\hat{r} > r_{\min} + \delta_r$ (note that δ_r can be chosen arbitrarily small in Algorithm 3.1). However, this contradicts the fact that $r_{\max}, r_{\max} \leq r_{\min} + \delta_r < \hat{r}$, has been proven in Algorithm 3.1 to be an infeasible solution of problem (15) since the required minimum power, p_R^* , is larger than the given constraint P_R in problem (15). Therefore, by contradiction, it follows that there does not exist such a feasible solution \hat{r} for problem (15). From the above discussions, it concludes that the feasible solution r_{\min} is at most δ_r lower than the optimal solution of problem (15). By letting $\delta_r \rightarrow 0$, convergence of Algorithm 3.1 is thus proved.

APPENDIX V PROOF OF THEOREM 3.2

Given \mathbf{X}^* , first we know that at least one of the two inequality constraints in (21) is active at the optimal point, i.e., we have either $\text{tr}(\mathbf{F}_1 \mathbf{X}^*) = 1$ or $\text{tr}(\mathbf{F}_2 \mathbf{X}^*) = 1$, or both. This fact can be proved by contradiction: If at \mathbf{X}^* both $\text{tr}(\mathbf{F}_1 \mathbf{X}^*) > 1$ and $\text{tr}(\mathbf{F}_2 \mathbf{X}^*) > 1$ hold, we could always find a t with $0 < t < 1$ such that $\mathbf{Y}^* = t \mathbf{X}^*$ and $\min(\text{tr}(\mathbf{F}_1 \mathbf{Y}^*), \text{tr}(\mathbf{F}_2 \mathbf{Y}^*)) = 1$. We could easily see that $\text{tr}(\mathbf{F}_0 \mathbf{Y}^*) < \text{tr}(\mathbf{F}_0 \mathbf{X}^*)$, which means that \mathbf{X}^* could not be the optimal solution, i.e., contradiction holds.

From now on, we assume w.l.o.g. that $\text{tr}(\mathbf{F}_1 \mathbf{X}^*) = 1$ such that we have $\text{tr}((\mathbf{F}_2 - \mathbf{F}_1) \mathbf{X}^*) \geq 0$. To facilitate the proof for Theorem 3.2, let us first give the following lemma, which is based on Lemma 1 given in [29], and the proof also follows a similar way to that in [29] (so it is skipped here).

Lemma 5.1: Given that $\text{tr}((\mathbf{F}_2 - \mathbf{F}_1) \mathbf{X}^*) \geq 0$, there exists a decomposition for \mathbf{X}^* such that

$$\mathbf{X}^* = \sum_{i=1}^r \mathbf{x}_i \mathbf{x}_i^T$$

and $\mathbf{x}_i^T (\mathbf{F}_2 - \mathbf{F}_1) \mathbf{x}_i \geq 0$ for all $i = 1, \dots, r$.

Based on the above lemma, let $y_{ij} = \mathbf{x}_j^T \mathbf{F}_i \mathbf{x}_j$, $i = 0, 1, 2$, and $j = 1, \dots, r$. Now consider the following linear program

$$\begin{aligned} \text{Min.}_{t_1, \dots, t_r} \quad & \sum_{j=1}^r y_{0j} t_j \\ \text{s.t.} \quad & \sum_{j=1}^r y_{1j} t_j \geq 1, \quad \sum_{j=1}^r y_{2j} t_j \geq 1 \\ & t_j \geq 0, \quad j = 1, \dots, r. \end{aligned} \quad (52)$$

We see that for any feasible set of t_1, \dots, t_r such that all the inequality constraints are satisfied, $\mathbf{X} = \sum_{j=1}^r t_j (\mathbf{x}_j \mathbf{x}_j^T)$ is a feasible solution for the SDP problem (21). As such, the minimum objective value of the above linear program is same as that of the SDP problem (21), and one such an optimal

point is $t_1 = \dots = t_r = 1$ (which corresponds to $\mathbf{X} = \sum_{j=1}^r t_j (\mathbf{x}_j \mathbf{x}_j^T) = \sum_{j=1}^r \mathbf{x}_j \mathbf{x}_j^T = \mathbf{X}^*$). Note that the optimal points may not be unique.

Furthermore, given that $\mathbf{x}_i^T (\mathbf{F}_2 - \mathbf{F}_1) \mathbf{x}_i \geq 0$ for all $i = 1, \dots, r$ from Lemma 5.1, we have $y_{2j} \geq y_{1j}$ for all j 's. Therefore, $\sum_{j=1}^r y_{1j} t_j \geq 1$ implies $\sum_{j=1}^r y_{2j} t_j \geq 1$, i.e., the second inequality constraint in (52) is redundant. Thus, (52) can be recast as

$$\begin{aligned} \text{Min. } & \sum_{j=1}^r y_{0j} t_j \\ \text{s.t. } & \sum_{j=1}^r y_{1j} t_j \geq 1 \\ & t_j \geq 0, \quad j = 1, \dots, r. \end{aligned} \quad (53)$$

When \mathbf{X}^* can be found for the SDP problem (21), it means that the optimal objective values for both the SDP problem (21) and the linear program problem (53) are bounded, which implies that (53) must have one basic optimal feasible solution, at which at least r inequality constraints are active (to define an optimal vertex point in the feasible region). Since we only have $r+1$ inequality constraints in (53), at most one t_j is positive. Actually, we have exactly one t_j positive; otherwise, all zero t_j 's could not be a feasible solution. At such a basic optimal feasible solution, if we have $t_k^* > 0$ and $t_j = 0, j \neq k$ with $1 \leq j, k \leq r$, we could infer that there exists an optimal rank-one solution for the SDP problem (21), which could be constructed as

$$\mathbf{X}^{**} = t_k^* (\mathbf{x}_k \mathbf{x}_k^T).$$

This completes the proof for Theorem 3.2.

At last, we present a routine to obtain an optimal rank-one solution for problem (20) from \mathbf{X}^* as follows:

- 1) Decompose \mathbf{X}^* in reference to $\mathbf{F}_2 - \mathbf{F}_1$ as in Lemma 5.1 (For detailed procedure, refer to the proof for Lemma 1 in [29]).
- 2) Construct the linear program problem as shown in (53), and solve one basic optimal feasible solution. Such an algorithm could be based on solving r parallel sub-problems, where at each sub-problem only one t_j is allowed to take non-zero values. Then the achieved minimum objective values from the r sub-problems are compared to find the global minimum solution.
- 3) Given the single optimal positive t_k^* , the rank-one optimal solution for both (20) and (21) is constructed as $\mathbf{X}^{**} = t_k^* (\mathbf{x}_k \mathbf{x}_k^T)$.

APPENDIX VI PROOF OF LEMMA 5.1

Let $a_{\text{MR}} = b_{\text{MR}} = \nu$ in \mathbf{A}_{MR} given by (22). We can then show the following equalities:

$$|\mathbf{h}_1^T \mathbf{A}_{\text{MR}} \mathbf{h}_2| = |\mathbf{h}_2^T \mathbf{A}_{\text{MR}} \mathbf{h}_1| = \nu \theta_1 \theta_2 (1 + \rho) \quad (54)$$

$$\|\mathbf{A}_{\text{MR}}^H \mathbf{h}_1^*\|^2 = \|\mathbf{A}_{\text{MR}} \mathbf{h}_1\|^2 = \nu^2 \theta_1^2 \theta_2 (1 + 3\rho) \quad (55)$$

$$\|\mathbf{A}_{\text{MR}}^H \mathbf{h}_2^*\|^2 = \|\mathbf{A}_{\text{MR}} \mathbf{h}_2\|^2 = \nu^2 \theta_1 \theta_2^2 (1 + 3\rho) \quad (56)$$

$$\text{tr}(\mathbf{A}_{\text{MR}} \mathbf{A}_{\text{MR}}^H) = 2\nu^2 \theta_1 \theta_2 (1 + \rho). \quad (57)$$

Let the relay transmit power p_R in (3) be equal to the maximum value P_R . Using the above equalities, from (3) it follows that

$$\nu^2 = \frac{P_R}{\theta_1 \theta_2 (1 + 3\rho) (\theta_1 p_1 + \theta_2 p_2) + 2\theta_1 \theta_2 (1 + \rho)}. \quad (58)$$

Substituting (58) into the above equalities, and from (6) and (7), the lower bound on the sum-rate given in (35) follows.

APPENDIX VII PROOF OF LEMMA 5.2

Let $a_{\text{ZF}} = b_{\text{ZF}} = \nu$ in \mathbf{A}_{ZF} given by (23). Denote $\mathbf{H}_{\text{UL}}^\dagger = [\mathbf{a}_1, \mathbf{a}_2]^T$. From (6) and (7), we can show that

$$\begin{aligned} R_{\text{LB}}^{\text{ZF}} &\geq \frac{1}{2} \log_2 \left(1 + \frac{\nu^2 p_2}{\|\mathbf{a}_2\|^2 \nu^2 + 1} \right) \\ &\quad + \frac{1}{2} \log_2 \left(1 + \frac{\nu^2 p_1}{\|\mathbf{a}_1\|^2 \nu^2 + 1} \right) \end{aligned} \quad (59)$$

$$\geq \log_2 \left(\frac{2p_1 p_2}{\|\mathbf{a}_2\|^2 p_1 + \|\mathbf{a}_1\|^2 p_2 + \frac{p_1 + p_2}{\nu^2}} \right) \quad (60)$$

where (60) is due to the Jensen's inequality (see, e.g., [30]) and the convexity of the function $f(x) = \log_2(1 + 1/x)$, $x \geq 0$ [25]. Let the relay transmit power p_R in (3) be equal to the maximum value P_R . Then, we obtain from (3) that

$$\nu^2 = \frac{P_R}{\|\mathbf{a}_2\|^2 p_1 + \|\mathbf{a}_1\|^2 p_2 + \text{tr}(\mathbf{H}_{\text{UL}}^\dagger (\mathbf{H}_{\text{UL}}^\dagger)^H (\mathbf{H}_{\text{DL}}^\dagger)^H \mathbf{H}_{\text{DL}}^\dagger)} \quad (61)$$

$$\geq \frac{P_R}{\|\mathbf{a}_2\|^2 p_1 + \|\mathbf{a}_1\|^2 p_2 + (\|\mathbf{a}_1\|^2 + \|\mathbf{a}_2\|^2)^2} \quad (62)$$

where (62) is due to the fact that $\text{tr}(\mathbf{XY}) \leq \text{tr}(\mathbf{X})\text{tr}(\mathbf{Y})$, if $\mathbf{X} \succeq 0$ and $\mathbf{Y} \succeq 0$ [31]. Using (62), the term inside $\log_2(\cdot)$ in (60) can be further lower-bounded by

$$\frac{2p_1 p_2}{\left(1 + \frac{p_1}{P_R} + \frac{p_2}{P_R} \right) (\|\mathbf{a}_2\|^2 p_1 + \|\mathbf{a}_1\|^2 p_2) + \frac{p_1 + p_2}{P_R} (\|\mathbf{a}_1\|^2 + \|\mathbf{a}_2\|^2)^2}.$$

Since it can be shown that $\|\mathbf{a}_1\|^2 + \|\mathbf{a}_2\|^2 = \frac{\theta_1 + \theta_2}{\theta_1 \theta_2 (1 - \rho)}$, substituting this equality into the above equation yields the lower bound on the sum-rate given in (36).

REFERENCES

- [1] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1204-1216, Jul. 2000.
- [2] Y. Wu, P. A. Chou, and S.-Y. Kung, "Information exchange in wireless networks with network coding and physical-layer broadcast," in *Proc. 39th Annual Conf. Inf. Sciences and Systems (CISS)*, Mar. 2005.
- [3] S. Katti, S. Gollakota, and D. Katabi, "Embracing wireless interference: Analog network coding," Computer Science and Artificial Intelligence Laboratory Technical Report, MIT-CSAIL-TR-2007-012, Feb. 2007.
- [4] S. Zhang, S.-C. Liew, and P. P. Lam, "Physical-layer network coding," in *Proc. IEEE MobiComm*, pp. 358-365, 2006.
- [5] M. P. W. K. Narayanan and A. Sprintson, "Joint physical layer coding and network coding for bi-directional relaying," in *Proc. 45th Annual Allerton Conference on Communication, Control and Computing*, pp. 254-259, Sep. 2007.
- [6] C. Schnurr, S. Stanczak, and T. J. Oechtering, "Coding theorems for the restricted half-duplex two-way relay channel with joint decoding," in *Proc. IEEE Int. Symp. Inf. Theory*, pp. 2688-2692, Jul. 2008.
- [7] S. Zhang and S.-C. Liew, "The capacity of two way relay channel." Available [Online] at <http://arxiv.org/abs/0804.3120>.

- [8] T. J. Oechtering, C. Schnurr, I. Bjelaković, and H. Boche, "Broadcast capacity region of two-phase bidirectional relaying," *IEEE Trans. Inf. Theory*, vol. 54, no.1 , pp. 454-458, Jan. 2008.
- [9] L.-L. Xie, "Network coding and random binning for multi-user channels," in *Proc. 10th Canadian Workshop on Information Theory*, pp. 85-88, Jun. 2007.
- [10] B. Rankov and A. Wittneben, "Spectral efficient signaling for half-duplex relay channels," in *Proc. 39th IEEE Asilomar Conf. on Signals, Systems and Computers*, Nov. 2005.
- [11] P. Larsson, N. Johansson, and K.-E. Sunell, "Coded bi-directional relaying," in *Proc. 63rd IEEE Veh. Technol. Conf. (VTC) -Spring*, pp. 851-855, May 2006.
- [12] P. Popovski and H. Yomo, "Bi-directional amplification of throughput in a wireless multi-hop network," in *Proc. 63rd IEEE Veh. Technol. Conf. (VTC) -Spring*, pp. 588-593, May 2006.
- [13] S. J. Kim, P. Mitran, and V. Tarokh, "Performance bounds for bi-directional coded cooperation protocols," *submitted to IEEE Trans. Inf. Theory*, 2007.
- [14] C. K. Ho, R. Zhang, and Y. C. Liang, "Two-way relaying over OFDM: optimized tone permutation and power allocation," in *Proc. IEEE ICC*, pp. 3908-3912, May 2008.
- [15] K. Jitvanichphaibool, R. Zhang, and Y. C. Liang, "Optimal resource allocation for two-way relay-assisted OFDMA," in *Proc. IEEE Global Commun. Conf. (Globecom)*, Dec. 2008.
- [16] I. Hammerström, M. Kuhn, C. Esli, J. Zhao, A. Wittneben, and G. Bauch, "MIMO two-way relaying with transmit CSI at the relay," in *Proc. IEEE Signal Proc. Adv. Wireless Comm. (SPAWC)*, Jun. 2007.
- [17] R. F. Wyrembski, T. J. Oechtering, I. Bjelaković., C. Schnurr, and H. Boche, "Capacity of Gaussian MIMO bidirectional broadcast channels," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2008.
- [18] T. Unger and A. Klein, "Linear transceive filters for relay stations with multiple antennas in the two-way relay channel," *16th IST Mobile and Wireless Communications Summit*, Budapest, Hungary, Jul. 2007.
- [19] T. Cui, F. F. Gao, T. Ho, and A. Nallanathan, "Distributed space-time coding for two-way wireless relay networks," in *Proc. IEEE ICC*, Beijing, China, May 2008.
- [20] X. Tang and Y. Hua, "Optimal design of non-regenerative MIMO wireless relays," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 1398-1407, Apr. 2007.
- [21] O. Munoz, J. Vidal, and A. Agustin, "Linear transceiver design in nonregenerative relays with channel state information," *IEEE Trans. Sig. Proces.*, vol. 55, pp. 2593-2604, Jun. 2007.
- [22] R. Zhang, C. C. Chai, and Y. C. Liang, "Joint beamforming and power control for multi-antenna relay broadcast channel with QoS constraints," *IEEE Trans. Sig. Proces.*, vol. 57, no.2, pp. 726-737, Feb. 2009.
- [23] Y. C. Liang and R. Zhang, "Optimal analogue relaying with multi-antennas for physical layer network coding," in *Proc. IEEE ICC*, pp. 3893-3897, May 2008.
- [24] R. Zhang, C. C. Chai, Y. C. Liang, and S. Cui, "On capacity region of two-way multi-antenna relay channel with analogue network coding," *to appear in IEEE ICC*, Jun. 2009.
- [25] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.
- [26] F. F. Gao, R. Zhang, and Y. C. Liang, "On channel estimation for amplify-and-forward two-way relay networks," in *Proc. IEEE Global Commun. Conf. (Globecom)*, Dec. 2008.
- [27] M. Mohseni, R. Zhang, and J. M. Cioffi, "Optimized transmission of fading multiple-access and broadcast channels with multiple antennas," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1627-1639, Aug. 2006.
- [28] Z.-Q. Luo and W. Yu, "An introduction to convex optimization for communications and signal processing," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1426-1438, Aug. 2006.
- [29] Y. Ye and S. Zhang, "New results on quadratic minimization," *SIAM J. Optim.*, vol. 14, no. 1, pp. 245-267, 2003.
- [30] T. Cover and J. Thomas, *Elements of Information Theory*, New York: Wiley, 1991.
- [31] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, 1985.